Università degli Studi di Trento

Department of Psychology and Cognitive Sciences (DiPSCO)

Doctoral school in Cognitive Sciences

Linguistic information and visual attention deployment: the influence of meaningful labels on the orienting of attention

By

Giulia Calignano

February 2020

ACKNOWLEDGEMENTS

In three years of passionate work, many are the people who contributed to help me get this important achievement. I start by thanking all the families and the participants that took part in my experiments and all the people I am going to forget in this short section but who made schools, workshops, lessons such fundamental steps on my way to the PhD.

Francesco Vespignani has been my lighthouse in Language science, along the whole doctoral path: he encouraged me and taught me how to be independent while collaborating with experts in linguistics, attention, cognition development and statistics fields. He inspired me not only as a great scientific supervisor but more importantly, as a wise friend.

Simone Sulpizio for the many insightful discussions, his energy and humour motivated me in many critical moments during the setup and the final editing of the entire project, he kept me going with outstanding sympathy and professionality.

Eloisa Valenza who have offered me illuminating discussions during the creative and hard process of scientific research in developmental cognition. She believed in my ideas and in my dedication without hiding any critical and insightful comments, whose I especially thank her for.

Rebecca Nako who outstandingly welcomed and supervised me at the Brain & Behaviour Lab and who taught me how to run N2pc studies with patience and joy. Martin Eimer a tireless mentor who offered me priceless theoretical and methodological advice during the fruitful London period.

Gary Lupyan e Gaia Scerif that dedicated their precious time to revise the first draft of this thesis by improving the whole work and by giving me amazing comments.

Alessandra Zappoli for our in-depth discussions about the ultimate sense of life in and out of academia and for sharing her knowledge about categorical perception of human speech. Sofia Russo for the long-standing collaboration that bind us like forever lab-life partner. Tobias Katus, Nick Berggren, Denise Baumeler and Alon Zivony for the many exciting scientific discussions at Birkbeck bar (and on the roof). Giulia Bini and Valentina Bologna who patiently helped me collecting infant data. Luca Menghini to be a great scientific and a loving partner who has never give up on me. Finally, thanks to Mirella, Antonio, Clara, Giuseppe and Maristella my family or I should say my tribe that pass me the need of intellectual freedom and beauty in anything I do.

ABSTRACT

The present work represents an endeavour towards the investigation of the linguistic-cognitive system under the lenses of classical questions in cognitive and language sciences, by using a multi-method and question oriented approach. The ambition is to move a step towards the investigation of the mutual contribution of perceptual and linguistic-mediated representations to the understanding of human behaviour. Chapter 1 will expose the theoretical framework and the goals this project was set to achieve: contributing to the theoretical reconcile of visual attention and language functions, from a developmental perspective. Chapter 2 will expose the possibility to rethink the linguistic function as penetrating human cognition in a top-down fashion, and specifically, its influence on template-guided search and disengagement of attention mechanisms. Concurrently, chapter 3 will expose the possibility to rethink the role of visual attention as a useful tool, necessary to the computation of meaning: attention will be introduced as a window to investigate the influence of language-mediated representation (spoken and written) on visuospatial mechanisms by means of ERPs and eye-tracking methodology. Finally, chapter 4 will report the rationale and the interpretation of seven original experimental investigations of the word (and sentence) effect on perceptual representation during visuospatial tasks, across infants and adults. The final discussion will try to reconcile the results of the presented studies with the theoretical and methodological issues raised in the first, second and third chapters in an integrated perspective of a linguistic-cognitive system.

INDEX

ABSTRACT 3

1. AIMS AND MOTIVATION OF THE THESIS 1

   1.1. Beyond Modularity: language-cognition interplay 2

   1.2. Ratio tyrannide liberare: the useful example of the instance theory of attention and memory (ITAM) 5

   1.3. The Label-Feedback Hypothesis (LFH) 8

   1.4. Beyond the Visual World Paradigm (VWP) 12

   1.5. Bibliography 14

2. LANGUAGE AS A COGNITIVE TOOL 25

   2.1. Word-Object pairs 28

   2.2. The sentence level..33

   2.3. Bibliografy 35

3. ORIENTING OF VISUAL ATTENTION: A WINDOW INTO MENTAL REPRESENTATION..41

   3.1. Disengagement of attention 43

   3.2. Looking times and pupil phasic response 46

   3.3. Target selection and identification 50

   3.4. The N2pc and SPCN: electrophysiological marker of template-guided search 52

   3.5. Bibliography 54

4. EMPIRICAL STUDIES 62

   4.1. Word-object pairs facilitate disengagement of attention in 12-month olds..62

   4.2. Disengagement of spatial attention as a window into word-object representation 93

4.3. Target representation during category-based visual search: is there a word advantage?..........122
4.4. Into the time course of sentence verification and visual target selection..........150
5. CONCLUSION..........183
5.1. Theoretical issues..........183
5.2. Methodological issues..........187
5.3. Strengths and further perspectives..........188
LIST OF FIGURES..........190
LIST OF TABLES..........195

1. AIMS AND MOTIVATION OF THE THESIS

It is commonly assumed that whatever the human intellectual capacity is, the faculty of language is essential to it (Chomsky, 2000). Yet, it remains an open question whether the generativity and the flexibility of the faculty of language can be generalised to other human cognitive functions (Gentner, 2010b, 2016; Karmiloff-Smith, 2009; Logan, 2002). On the one side, specific domain theories expect language and visual attention to be multi-component, universal and generative functions, that eventually interact by developing solutions to specific tasks (Myachykov & Posner, 2005b; Posner, Snyder, & Solso, 2004). On the other side, general domain theories expect universal and generative faculties to explain behavioural outcomes across tasks (Hauser & Watumull, 2017). Somewhere in the middle, converging evidence from different approaches to cognitive development, like studies on early attention development (Colombo & Posner, 2004), language acquisition (Dispaldro et al., 2013; Saffran, 2002; Xu & Tenenbaum, 2007), early categorisation of word-object pairs (Ferry, Hespos, & Waxman, 2010; Pomper & Saffran, 2018), and on the effect of word learning on attention deployment (Zamuner, Fais, & Werker, 2014) suggests that the influence exerted by linguistic and perceptual representations on attention and linguistic mechanisms is more likely to share variance in a contextual and time-dependent fashion, see Figure 1.1. (Karmiloff-Smith, 2000).

In this framework, the present project tried to address at least two issues. One the one hand, (1) questions that do challenge theoretical issues, like ‘do infants and adults make similar use of the referential link of word-object pairs during attention deployment?’ (see chapter 4.1.1 and 4.1.2), or ‘does linguistic information trigger perceptual representation, and vice versa?’ (see chapter 4.2.1), and if it is the case, ‘do linguistic structures shape pictorial-guided visual search?’ (chapter 4.2.2).

Seven experimental investigations tried to disentangle these questions by stressing the cognitive and the language system in tasks set to investigate whether and how these two functions overlap (Perry & Samuelson, 2011).
On the other hand, this work has been oriented by (2) questions that challenge methodological issues, like 'do converging evidence from a multi-method approach offer a view on the whole cognitive system, whereas a unique-method approach may unambiguously replicate itself, but it offers a small view on each mechanism, separately?'. The challenge of this kind of methodological issues may contribute to reconcile a definition of replication in psychology (Maxwell, Lau, & Howard, 2015).

1.1. Beyond Modularity: language-cognition interplay

During the last fifty-seven years of 'cognitive-revolution' (Chomsky, 1992, 2012) the number of experimental evidences end methodological advancements have exponentially increased. Nevertheless, the field has been tacitly dominated by latent reductionism (Fairweather, Brizzolara, Tabossi, & Umiltà, 1982; Gallese, 2009) that has made it hard so far, to reconcile the growing amount of behavioural data coming from different approaches - a kind of Jekyll and Hyde. On one very extreme, it is assumed that the investigation of a single cognitive function independently from other cognitive functions, cannot lead to the understanding of the sizeable unexplained variance of human behaviour: because the cognitive system is a flexible system that is affected by the time-course of events. On the very other extreme, psychology, linguistics, neuroscience, and philosophy persists in proposing models of the cognitive system that tacitly assume a hierarchy of separate functions shaped on a static adult brain.

At least three issues have implicitly dominated this framework: *innateness* — that implies that the infant brain is already equipped with adult cognitive functions; *localization* — that reflects the empirical attempt of finding the correlates of innate functions in specific networks of the brain; and *domain specificity* — that expects specialised core of knowledge to be linearly mapped into separate functions (Bates, 1999; Karmiloff-Smith, 1999; Karmiloff-Smith, Scerif, & Ansari, 2003). Otherwise, the framework expects general-domain mechanisms to orchestrate information processing independently from the specific domain of knowledge (for a critical discussion see Bates, 1999; KarmiloffSmith, 2000).

A third way has been indicated by several authors who have carried out a Copernican revolution by reconciling the dominating adult-centric, highly modular and domain-specialized models with a development-centric, time-dependent understanding of the cognitive functions (Thelen & Smith, 1996; Karmiloff-Smith, 2018). The adult cognitive system is defined as the product of several small changes that prompt complex outcomes from the prenatal period and throughout development; to such extent, some authors have proposed that the develop*ing* brain might not fully overlap the develop*ed* brain in terms of structure and functions (Karmiloff-Smith, 2015). The Neuroconstructivist framework has proposed to overcome the domain-specificity-vs-generality dualism, by indicating a third explanatory way in "domain-relevant mechanisms". The domain-relevant mechanisms reflect the interaction between biological and environmental constraints and gradually specialise along the developmental pathway. This framework does not exclude any mechanisms to be domain-specific, but it expects specificity to be the results of development pathways and task demands, since the prenatal period (Cornish, Scerif, & Karmiloff-Smith, 2007; Karmiloff-Smith, Casey, Massand, Tomalski, & Thomas, 2014). The domain-relevant mechanisms opened the way to investigate the flexibility of the whole cognitive system by expecting dramatic changes in the brain structure and functions along time (Karmiloff-Smith & Inhelder, 1974).

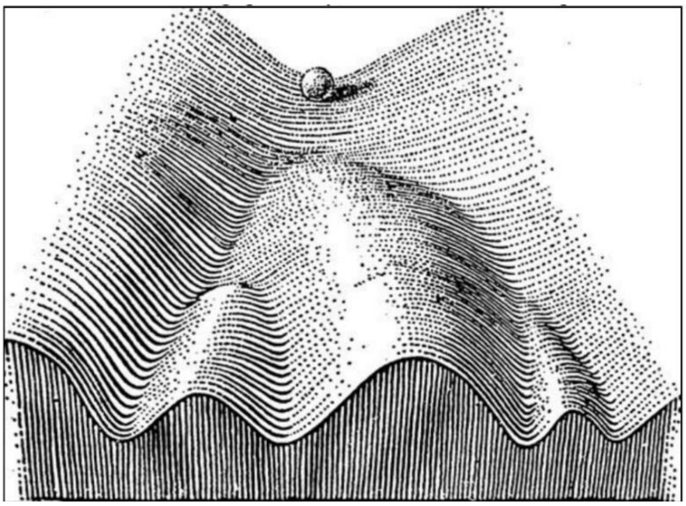


Figure 1. The Waddington's Epigenetic Landscape in which the ball represents the individual, at the very beginning of development. The outcome depends on the interaction between the biological constraints carried by the individual (e.g. balls do roll) and the physical constraints on the situated environment (e.g. shape of the landscape).

For clear-cut reasons, language acquisition and computation have been placed at the centre of this debate (Karmiloff-Smith, 2015; Kover, McCary, Ingram, Hatton, & Roberts, 2015; Thomas & Karmiloff-Smith, 2003). Innateness (Ganger & WOLD, 1998; Hauser, Chomsky, & Fitch, 2002), localization (Carreiras, Armstrong, Perea, & Frost, 2014; Price, 2012) and domain-specificity (Shafto, Conway, Field, & Houston, 2012) are milestone issues in language sciences. Despite domain-relevant mechanisms ascribed to the brain activity are expected to interact multi-directionally rather than relying on deterministic gene-behaviour rules (see Figure 1, the Epigenetic Landscape Weddington, 1957; for some recent evidence embracing this framework see Mento, Scerif, Granziol, Franzoi, & Lanfranchi, 2019; Campos, Nieto, & Núñez, 2019; Quadrelli & Turati, 2016), the Neuroconstructivist framework has been mainly confined in developmental cognition research (Bates et al., 1998; Resende, 2019). This thesis investigates the extent by which this epigenetic framework may be extended to models of adult behaviour (see also Manning, Scerif, & Norcia, 2018) by expecting pictorial and linguistic-mediated representation to influence attention and linguistic mechanisms depending on the task specificity rather than on separate and innate modules (Brooks, Kempe, & Deák, 2014).

1.2. Ratio tyrannide liberare: the useful example of the instance theory of attention and memory (ITAM)

The dominant view has puzzled the investigation of every-day human behaviour in a fashion that the same act of cognition has been defined and operationalised in several ways that do not exclude each other, yet do not integrate across disciplines. Let us consider an example that can be easily understood. Imagine Mary is travelling from her place to a foreign country she does not master the language of. After a long way trip, Mary lands on the destination and well, she is starving. Eventually, she can find a little restaurant or, at least, a building that nicely mirrors her canonical representation of restaurants. In such a specific case, Mary expects to see a restaurant sign, people sitting on laid tables calling the waiter time to time, and to smell hot food coming from inside a kitchen. A student of attention would be interested in how Mary does (or does not) pick the restaurant among the group of buildings. A student of semantics would be interested in how Mary does (or does not) categorise the meaning of restaurant according to her knowledge about the real-world entities such as food, tables and shop-signs. The two students' different approach to understand the same act of cognition do reflect the dominant framework: "different researchers focus on the details of different parts of cognition, hoping that their work will interface with the rest"(Logan, 2002). Unfortunately, the findings of each investigation will hardly interface with the rest.

This chapter aimed to provide a scenario in which, according to the previous example, a student (in this case me) is interested in how Mary picks the restaurant among other buildings while dealing with perceptual and linguistic representations. By embracing a domain-relevant perspective, the student, in this case, is interested in studying how different mechanisms lead to similar outcomes and how different outcomes be explained by common mechanisms, across time. That is, attention may begin with perception for both infant and adult brains. Categorisation may begin when a visual target is found and end with a label, in adults (Zamuner, Fais, & Werker, 2014). Whereas it may begin when a label is found and end with a visual target in infants (Waxman & Guasti, 2009). Thus, in line with the general intentions of Logan (2002) this work opposites a *divide et impera* approach in favour of a *ratio tyrannide liberare* perspective that views "*the simple act of cognition as a single phenomenon and interprets attention, categorisation as different perspectives on the same simple act. To attend is to categorise; to categorise is to remember; to remember is to attend*". Logan (2002) among others proposed the instance theory of attention and memory (ITAM) that encompassed some classical models of attention (attention branch including at least three dimensions of attention sub-components i.e., attention to objects, attention to categories, and attention to dimensions; see figure 2) and some classical model of memory (memory branch) that involves linguistic mediated models (meaning retrieval).

This theory reconciles linguistic and conceptual representations during attention and memory interface by crossing the borders of domain-specificity to explain Mary's behaviour. It assumes a reference frame computation (context model instance theory) as a necessary step for orienting attention to occur in classical cueing tasks, in which participants are instructed to select a target specified by a cue. Logan (2002) reference frames, like the spatial reference frames (e.g., next to, above, below) are defined as mechanisms of attention just like spotlights and spatial indices, which computation flexibility mirrors traditional attentional mechanisms. The spatial frame computation is a good candidate to be a relevant domain mechanism mediated by linguistic and conceptual representations.

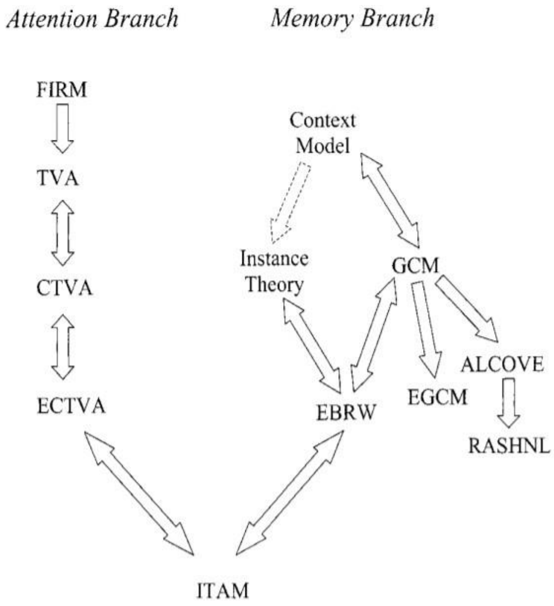


Figure 2. Picture from Logan (2002), A family tree expressing the relations between the ancestors of the attention branch and those of the memory branch. Theories connected by bidirectional arrows can be made equivalent to each other, through ITAM. FIRM = fixed-capacity independent race model; TVA = theory of visual attention; CODE = COntour Detector theory of visual attention; ECTVA = executive control of TVA; GCM = generalized context model; EBRW = exemplar-based random walk model; EGCM = extended generalized context model; ALCOVE = attention learning COVEring theory; RASHNL = rapid attention Shifts 'N' Learning theory; ITAM = instance theory of attention and memory.

Thus, the analysis of the linguistic interference during typical visuospatial controlled experiment may help to investigate the nature of the representation and the mechanisms guiding attention deployment. Investigating the control exerted by the reference frame prompted by language-mediated representations means taking charge of an overlapping language/cognitive system. Here, I focused on orienting of visual attention and language (single words and sentences), by trying to integrate different approaches oriented by the question of whether domain-relevant mechanisms (reference frame computation) are similarly affected by pictorial and linguistics representations in specific circumstances.

1.3. The Label-Feedback Hypothesis (LFH)

Language can be understood as participating in the cognitive system in a top-down fashion. Here, top-down refers to information about a target identity where this information is not present in the visual stimulus (Wolfe, Horowitz, Kenner, Hyle, & Vasan, 2004a). According to this, the Label-feedback Hypothesis (Lupyan, Rakison, & McClelland, 2007) suggests that any linguistic unit is categorical in nature and points to (at least) a category of distinctive features (Coco et al., 2014; Konkle, Brady, & Alvarez, 2010; Maxfield, Stalder, & Zelinsky, 2014). Thus, labels act as categorical top-down information that triggers pictorial representation and in turn, facilitates perceptual processing (see Figure 3, Lupyan & Thompson-schill, 2012).

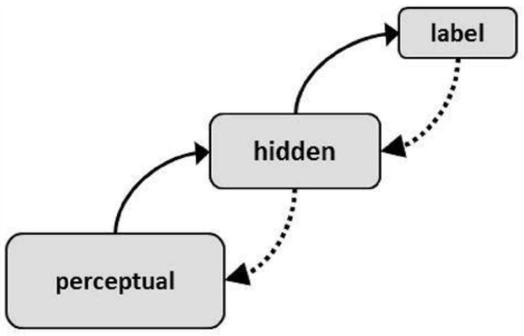


Figure 3. The LFH model (Lupyan, 2012). A schematic of a neural network architecture for exploring on-line effects of labels on perceptual representations.

Indeed, the ability to recognise sounds associated to real objects can be more context-dependent compared to word recognition, which representation retrieval stays more stable across contexts (Edmiston & Lupyan, 2015; Kristensen, Wang, Petersson, & Hagoort, 2013; Peng et al., 2010). Boutonnet and Lupyan (2015), showed that the presentation of the valid spoken word cue (e.g. dog) predicted faster RT (for valid responses) compared to the presentation of the corresponding sound (e.g. a dog bark), in a visual recognition task: to identify a visual target (e.g. a dog) a facilitatory effect was driven by on-line label presentation. Critically, this effect emerged by comparing words and sounds that do have a close semantic link to the target. Thus, words seem to preferentially trigger template of category items (dogs) during target identification, compared to general audio and video information *per se*. In a further study with the Continuous Flash Suppression (CFS), a technique – that takes advantage of the binocular competition induced by the presence of different stimuli per eye (e.g. a drawn apple to the right eye and visual noise to the left eye). Lupyan and Ward (2013) used the CFS to prevent the target recognition. They found that when a pre-cue precedes a noise stimulus, participants detected the target (e.g. an apple) more accurately compared to an invalid pre-cue (e.g., banana) and no pre-cue conditions i.e., silence. Critically, neither the target familiarity nor the complexity of the word structure seemed to play a role in the on-line label facilitatory effect, as shown by using novel objects and labels never encountered by the participants before the controlled experiment in the laboratory (Lupyan & Thompson-schill, 2012).

Although evidence suggests an on-line linguistic facilitatory effect during object recognition, the effects of the labels during guided search are controversial. On the one hand, evidence coming from template-guided search shows that word cues are less effective compared to pictorial cues (Nako, Wu, Smith J., & Eimer, 2014; Wolfe, Horowitz, Kenner, Hyle, & Vasan, 2004b). The main explanation relies on the fact that the effectiveness of a cue depends on its ability to specify the information needed for the search (Wolfe & Horowitz, 2004). Indeed, the attentional template – the mental representation of the target that recollects its relevant features in WM (Eimer, 1996, 2014) – based on an pictorial cue that exactly match the target is established immediately (less than 50ms from cue onset), as shown by the RTs facilitation in search tasks triggered by pictorial compared to word cue, at least with short interval (SOA less than 800ms, Wolfe et al., 2004). Given that the more definite the target template the more accurate the target selection can be, it is not surprising that the exact reproduction of the target leads to faster RTs and a more accurate visual search with respect to a word cue, which specify a broad perceptual and semantic category, e.g., apple, fruit, food (Nako et al., 2014; Schmidt & Zelinsky, 2009; Vickery, King, Treisman, Wolfe, & Duncan, 2018; Wolfe et al., 2004a).

On the other hand, some picture cues are more effective than others. Indeed, if a picture cue is smaller or bigger or rotated or occluded or negated (Arita, Carlisle, & Woodman, 2012), this facilitation effect is not as large as the exact pictorial cue (Vickery et al., 2005). In addition, when a pictorial cue is not an exact reproduction of the target, the pictorial superiority effect disappears (Wolfe et al., 2004). Furthermore, labels of canonical objects (i.e., banana) exhibit higher frequency of co-occurrence and consequently, prompt more efficient templates compared to labels of non-canonical objects (i.e., bags) during visual search task by suggesting that words prompt canonical perceptual representations. These evidences suggest that linguistic-mediate representation might play a role during template-guided search. According to Soto & Humphreys (2007), the possibility that a perceptual representation is retrieved through the linguistic cue is an insufficient explanation. Indeed, they showed that articulatory suppression could block the visual attentional capture guided by pictorial cue in WM, during a visual search task. The authors suggest that in humans, any retrieval of an image is not efficient in the absence of verbalisation. Previously, Potter, Kroll, and Yachzel (1986) findings showed that previous verbal description of a visual target could enhance target identification to the same extent of a visual cue, even when the verbal description does not provide any information about the target colour or shape (rebus sentences). Some interesting insights come from studies that investigates the linguistic mediated representation employing the Visual World Paradigm, in which participants are asked to carefully listen to a spoken sentence while freely looking at pictures presented on the screen (see Figure 4, Huettig, Olivers, & Hartsuiker, 2011; Huettig, Rommers, & Meyer, 2011).

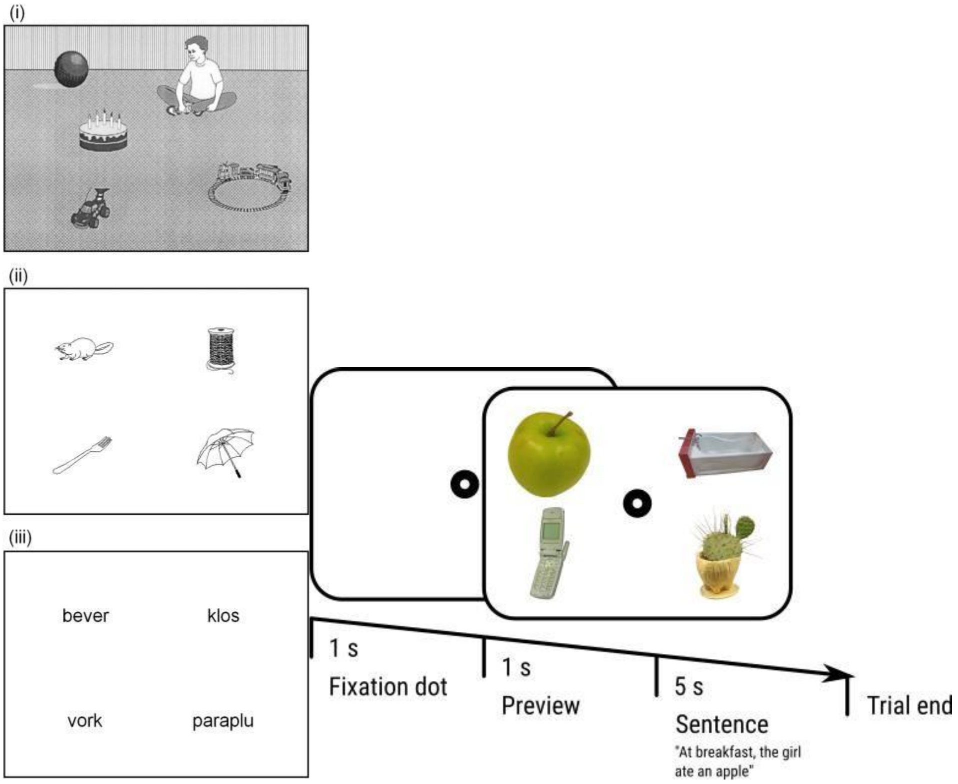


Figure 4. The Visual Word Paradigm by, On the left panel, the typical visual displays in the Visual World Paradigm (Huettig et al., 2011b). (i) an example of a semirealistic scene (participants either heard "The boy will eat the cake" or "The boy will move the cake") (Altmann and Kamide, 1999). (ii) an example of a four object display and (iii) is an example of a printed word display (Huettig and McQueen, 2007). On the right panel, an example of a Full Match trial in a VWP, because the target object (the apple) is directly mentioned in the spoken sentence. Stimuli taken from the BOSS stimuli (Brodier et al., 2010) form the OpenSesame documentation.

Such a task has allowed to show that visual attention is guided in a top-down fashion by semantically related items (e.g., a lock while listening to the word key; Moores, Laiti, & Chelazzi, 2003); or a trumpet at the spoken word piano (Huettig & Altmann, 2011) or phonological neighbours (e.g., word labels rhyme, begin with the same sound or are similar-sounding (Allopenna, Magnuson, & Tanenhaus, 1998). This kind of semantic bias has been found also when the cue is the exact target picture (Walenchok, Hout, & Goldinger, 2016). Accordingly, with the IATM (Logan, 2002) mentioned in the previous paragraph, such evidence allows us to conclude that rapid processing of the visual scene requires the computation of a semantic frame at a more conceptual level, mediated by linguistic and pictorial representations. Indeed, visual imagery cannot fully account for these findings, since in general, an explanation based on the visual similarity is less informative about semantic biases, that likely depends on conceptual hierarchy (Maxfield & Zelinsky, 2012).

1.4. Beyond the Visual World Paradigm (VWP)

The VWP has been introduced in psycholinguistics as a tool to bridge language (comprehension and production) and visual processing, and it had been a significant impact on the field by contributing to the debate about structural (i.e. early syntactical biases) and constraint-based models (i.e. early semantic biases) (Huettig, Rommers, & Meyer, 2011b).

Like any paradigm in cognitive research, the VWP carries outs limits to the kind of questions it can help answer. For instance, it is not evident whether the measure it offers accounts for the effect of linguistic representation on visual attention mechanisms, or rather the perceptual representation triggered by the visual stimuli plays a significant role in explaining language-mediated eye movements (for a detailed discussion see Huettig, Olivers, & Hartsuiker, 2011). Some studies aimed to disentangle this issue and showed that presenting a blank screen after the search array do rise the likelihood of the participants' tendency to re-fixate the regions on the blank screen that were previously occupied by the mentioned objects (see Altmann, 2004; Altmann & Kamide, 2007). This critical result has been explained in terms of linguistic-mediated effect prompted off-line by linguistic saliency. However, an explanation not mutually exclusive with the previous one can be found on after-effects lead by perceptual saliency of visual objects guiding the linguistic verification (Lüdtke, Friedrich, De Filippis, & Kaup, 2008). Chapter 4.2 aims to offer some methodological alternatives to disentangle these explanations by investigating visual attention deployment during word (4.2.1) and sentence (4.2.2.) verification during active visual search tasks that allow to isolate target feature selection depending on cue specificity, and specifically in Category-guided search that allow to stress the effect of linguistic mediated representation, and in a Sentence Picture Verification that allows to controlling for semantic and syntactic bias, in a pragmatic context.

A further critical issue faced by the VWP is related to familiarity bias. The VWP requires visual objects to be real-objects or real-actions that allow for manipulating language complexity in controlled experiments and, accordingly, measuring language comprehension and production ability. However, although participants share a significant portion of variance in terms of frequency of labels and visual stimuli exposure, each of them brings an critical portion of biases that guides resource allocation (Perry & Samuelson, 2011). For example, the age of acquisition and the frequency by which word-object pairs co-occur in the own environment, as well as the affective meaning associated with everyday objects varies as a function of several factors. Then, introducing in the laboratory a Training-Test paradigm with novel objects and unknown labels allows to test language effects on visual and attention mechanisms by adding a degree of freedom and by limiting familiarity confounds (Lupyan & Thompson-schill, 2012; Sulpizio & Mcqueen, 2012). Moreover, it allows to investigate the effect of linguistic and pictorial representations on attention mechanisms in infant ad adult population, in similar tasks (see Chapter 4.1.1 and 4.1.2). Finally, psycholinguistics research has focused its investigation mainly on active processes like language verification and cueing tasks with words, trying to disentangle the strength by which the linguistic representations affect human behaviour. Nevertheless, the investigation of top-down influence on visuospatial deployment that does not necessarily require identification has not been widely integrated in visual attention models.

1.5. Bibliography

Arita, J. T., Carlisle, N. B., & Woodman, G. F. (2012). Templates for rejection: configuring attention to ignore task-irrelevant features. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(3), 580.

Fairweather, H., Brizzolara, D., Tabossi, P., & Umiltà, C. (1982). Functional cerebral lateralisation; dichotomy or plurality? *Cortex*, *18*(1), 51–65.

Gallese, V. (2009). Motor abstraction: A neuroscientific account of how action goals and intentions are mapped and understood. *Psychological Research PRPF*, *73*(4), 486–498.

Huettig, F., Olivers, C. N. L., & Hartsuiker, R. J. (2011). Acta Psychologica Looking , language , and memory : Bridging research from the visual world and visual search paradigms. *ACTPSY*, *137*(2), 138–150. https://doi.org/10.1016/j.actpsy.2010.07.013

Karmiloff-Smith, A. (2015). An alternative to domain-general or domain-specific frameworks for theorizing about human evolution and ontogenesis. *AIMS Neuroscience*, *2*(2), 91–104. https://doi.org/10.3934/Neuroscience.2015.2.91

Kaup, B., Lüdtke, J., & Zwaan, R. A. (2006). Processing negated sentences with contradictory predicates: Is a door that is not open mentally closed? *Journal of Pragmatics*, *38*(7), 1033–1050. https://doi.org/10.1016/j.pragma.2005.09.012

Lewkowicz, D. J. (2014). *Perception of Auditory-Visual Temporal Synchrony in Human Infants Perception of Auditory-Visual Temporal Synchrony in Human Infants*. (November 1996). https://doi.org/10.1037/0096-1523.22.5.1094

Lewkowicz, D. J., Schmuckler, M. A., & Mangalindan, D. M. J. (2018). Learning of hierarchical serial patterns emerges in infancy. *Developmental Psychobiology*, *60*(3), 243–255. https://doi.org/10.1002/dev.21614

Logan, G. D. (2002). An instance theory of attention and memory. *Psychological Review*, *109*(2), 376.

Lüdtke, J., Friedrich, C. K., De Filippis, M., & Kaup, B. (2008). Event-related potential correlates of negation in a sentence-picture verification paradigm. *Journal of Cognitive Neuroscience*, *20*(8), 1355–1370.https://doi.org/10.1162/jocn.2008.20093

Matusz, P. J., Merkley, R., Faure, M., & Scerif, G. (2019). Expert attention: Attentional allocation depends on the differential development of multisensory number representations. *Cognition*, *186*, 171–177.

Maxwell, S. E., Lau, M. Y., & Howard, G. S. (2015). Is psychology suffering from a replication crisis? What does "failure to replicate" really mean? *American Psychologist*, *70*(6), 487.

Neil, P. A., Chee-Ruiter, C., Scheier, C., Lewkowicz, D. J., & Shimojo, S. (2006). Development of multisensory spatial integration and perception in humans. *Developmental Science*, *9*(5), 454–464. https://doi.org/10.1111/j.14677687.2006.00512.x

Perry, L. K., & Samuelson, L. K. (2011). The shape of the vocabulary predicts the shape of the bias. *Frontiers in Psychology*, *2*, 345.

Robinson, C. W., Best, C. A., Deng, W. S., & Sloutsky, V. M. (2012). *The role of words in cognitive tasks : what , when , and how ? 3*(April), 1–8. https://doi.org/10.3389/fpsyg.2012.00095

Robinson, C. W., & Sloutsky, V. M. (2004). Auditory dominance and its change in the course of development. *Child Development*, *75*(5), 1387–1401.

Scappini, M., Vespignani, F., & Delfitto, D. (2015). *On the non-incremental processing of negation: A pragmatically licensed sentence-verification study with Italian adults*.

Sulpizio, S., & Mcqueen, J. M. (2012). Italians use abstract knowledge about lexical stress during spoken-word recognition. *Journal of Memory and Language*, *66*(1), 177–193. https://doi.org/10.1016/j.jml.2011.08.001

Tian, Y., Breheny, R., & Ferguson, H. J. (2010). Why we simulate negated information: A dynamic pragmatic account. *The Quarterly Journal of Experimental Psychology*.

Tian, Y., Ferguson, H., & Breheny, R. (2016). *Processing negation without context why and when we represent the positive argument. 3798*. ttps://doi.org/10.1080/23273798.2016.1140214

Waxman, S. R., & Gelman, S. A. (2009). *Early word-learning entails reference , not merely associations*. (May), 258–263. https://doi.org/10.1016/j.tics.2009.03.006

Waxman, S. R., & Teresa Guasti, M. (2009). Nouns, Adjectives, and the Acquisition of Meaning: New Evidence from Italian-Acquiring Children. *Language Learning and Development*, *5*(1), 50–68. https://doi.org/10.1080/15475440802347528

Zamuner, T. S., Fais, L., & Werker, J. F. (2014). *Infants track word forms in early word – object associations*. *4*, 481–491. https://doi.org/10.1111/desc.12149

Bates, E. (1999). Language and the infant brain. *Journal of Communication*

Disorders, *32*(4), 195–205.

Bates, E., Elman, J., Johnson, M., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1998). Innateness and emergentism. *A Companion to Cognitive Science*, 590–601.

Boutonnet, X. B., & Lupyan, X. G. (2015). *Words Jump-Start Vision : A Label Advantage in Object Recognition*. *35*(25), 9329–9335. https://doi.org/10.1523/JNEUROSCI.5111-14.2015

Brooks, P., Kempe, V., & Deák, G. O. (2014). Interrelationship of Language and Cognitive Development (Overview). *Encyclopedia of Language Development*, 284–291. https://doi.org/10.4135/9781483346441.n91

Campos, R., Nieto, C., & Núñez, M. (2019). Research domain criteria from neuroconstructivism: A developmental view on mental disorders. *Wiley Interdisciplinary Reviews: Cognitive Science*, *10*(3), e1491.

Carreiras, M., Armstrong, B. C., Perea, M., & Frost, R. (2014). The what , when , where , and how of visual word recognition. *Trends in Cognitive Sciences*, *18*(2), 90–98. https://doi.org/10.1016/j.tics.2013.11.005

Chomsky, N. (2000). *The architecture of language*.

Coco, M. I., Keller, F., T., B. G., R., C., J., H., K., R., … N., V. W. (2014). Classification of visual and linguistic tasks using eye-movement features. *Journal of Vision*, *14*(3), 11. https://doi.org/10.1167/14.3.11

Colombo, J., & Posner, M. I. (2004). Visual Attention in Infancy: Process and Product in Early Cognitive Development. *Cognitive Neuroscience of Attention.*, (JANUARY 2004), 329–341.

Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology*.

Cornish, K., Scerif, G., & Karmiloff-Smith, A. (2007). Tracing syndrome-specific trajectories of attention across the lifespan. *Cortex*, *43*(6), 672–685.

Dispaldro, M., Leonard, L. B., Corradi, N., Ruffino, M., Bronte, T., & Facoetti, A. (2013). Visual attentional engagement deficits in children with Specific Language Impairment and their role in real-time language processing. *Cortex*, *49*(8), 2126–2139. https://doi.org/10.1016/j.cortex.2012.09.012

Edmiston, P., & Lupyan, G. (2015). What makes words special ? Words as unmotivated cues. *Cognition*, *143*, 93–100. https://doi.org/10.1016/j.cognition.2015.06.008

Eimer, M. (1996). The N2pc component as an indicator of attentional selectivity. *Electroencephalography and Clinical Neurophysiology*, *99*(3), 225–234.

Eimer, M. (2014). The neural basis of attentional control in visual search. *Trends in Cognitive Sciences*, *18*(10), 526–535.

Elman, J. L., Bates, E., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). Rethinking innateness: Connectionism in a developmental framework. *Cambridge, MA: MIT Press. Embick, D., Marantz, A., Miyashita, Y., O'Neil W., Sakai, KL (2000). A Syntactic Specialization for Broca's Area. Proceedings of the National Academy of Sciences*, *97*(11), 6150–6154.

Ferry, A. L., Hespos, S. J., & Waxman, S. R. (2010). Categorization in 3- and 4month-old infants: An advantage of words over tones. *Child Development*, *81*(2), 472–479. https://doi.org/10.1111/j.1467-8624.2009.01408.x

Ganger, J., & WOLD, K. S. (1998). Innateness, evolution, and genetics of language. *Human Biology*, 199–213.

Gentner, D. (2010a). Bootstrapping the mind: Analogical processes and symbol systems. *Cognitive Science*, *34*(5), 752–775.

Gentner, D. (2010b). Psychology in cognitive science: 1978–2038. *Topics in Cognitive Science*, *2*(3), 328–344.

Gentner, D. (2016). Language as cognitive tool kit: How language supports relational thought. *American Psychologist*, *71*(8), 650.

Gentner, D., & Hoyos, C. (2017). *Analogy and Abstraction*. *9*, 672–693. https://doi.org/10.1111/tops.12278

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The Faculty of Language: What Is It, Who Has It, and How Did It Evolve? *Science*, *298*(5598), 1569–1579. https://doi.org/10.1126/science.298.5598.1569

Horst, J. S., & Hout, M. C. (2016). The Novel Object and Unusual Name (NOUN) Database: A collection of novel images for use in experimental research. *Behavior Research Methods*, *48*(4), 1393–1409.

Huettig, F., & Altmann, G. T. M. (2011). Looking at anything that is green when hearing "frog": How object surface colour and stored object colour knowledge influence language-mediated overt attention. *Quarterly Journal of Experimental Psychology*, *64*(1), 122–145. https://doi.org/10.1080/17470218.2010.481474

Huettig, F., Mishra, R. K., & Olivers, C. N. L. (2012). Mechanisms and representations of language-mediated visual attention. *Frontiers in Psychology*, *3*(JAN), 1–11. https://doi.org/10.3389/fpsyg.2011.00394

Huettig, F., Olivers, C. N. L., & Hartsuiker, R. J. (2011). Acta Psychologica Looking , language , and memory : Bridging research from the visual world and visual search paradigms. *ACTPSY*, *137*(2), 138–150. https://doi.org/10.1016/j.actpsy.2010.07.013

Huettig, F., Rommers, J., & Meyer, A. S. (2011a). Acta Psychologica Using the visual world paradigm to study language processing : A review and critical evaluation. *ACTPSY*, *137*(2), 151–171. https://doi.org/10.1016/j.actpsy.2010.11.003

Huettig, F., Rommers, J., & Meyer, A. S. (2011b). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, *137*(2), 151–171.

Karmiloff-Smith, A. (1999). Taking development seriously. *Human Development*, *42*(6), 325–327.

Karmiloff-Smith, A. (2000). Why babies' brains are not Swiss army knives. *Alas, Poor Darwin: Arguments against Evolutionary Psychology*, 144–156.

Karmiloff-Smith, A. (2009). Nativism versus neuroconstructivism: rethinking the study of developmental disorders. *Developmental Psychology*, *45*(1), 56.

Karmiloff-Smith, A. (2013). Challenging the use of adult neuropsychological models for explaining neurodevelopmental disorders: Develop ed versus develop ing brains. *The Quarterly Journal of Experimental Psychology*, *66*(1), 1–14.

Karmiloff-Smith, A. (2015). An alternative to domain-general or domain-specific frameworks for theorizing about human evolution and ontogenesis. *AIMS Neuroscience*, *2*(2), 91–104. https://doi.org/10.3934/Neuroscience.2015.2.91

Karmiloff-Smith, A. (2018). Development itself is the key to understanding developmental disorders. In *Thinking Developmentally from Constructivism to Neuroconstructivism* (pp. 97–117). Routledge.

Karmiloff-Smith, A., Casey, B. J., Massand, E., Tomalski, P., & Thomas, M. S. C. (2014). Environmental and genetic influences on neurocognitive development: The importance of multiple methodologies and timedependent intervention. *Clinical Psychological Science*, *2*(5), 628–637.

Karmiloff-Smith, A., & Inhelder, B. (1974). If you want to get ahead, get a theory. *Cognition*, *3*(3), 195–212.

Karmiloff-Smith, A., Scerif, G., & Ansari, D. (2003). Double dissociations in developmental disorders? Theoretically misconceived, empirically dubious. *Cortex*, *39*(1), 161–163.

Konkle, T., Brady, T. F., & Alvarez, G. A. (2010). *Conceptual Distinctiveness Supports Detailed Visual Long-Term Memory for Real-World Objects*. *139*(3), 558–578. https://doi.org/10.1037/a0019165

Kover, S. T., McCary, L. M., Ingram, A. M., Hatton, D. D., & Roberts, J. E. (2015). Language development in infants and toddlers with fragile X syndrome: Change over time and the role of attention. *American Journal on Intellectual and Developmental Disabilities*, *120*(2), 125–144.

Kristensen, L. B., Wang, L., Petersson, K. M., & Hagoort, P. (2013). *The Interface Between Language and Attention : Prosodic Focus Marking Recruits a General Attention Network in Spoken Language Comprehension*. (August), 1836–1848. https://doi.org/10.1093/cercor/bhs164

Lerner, B. D. (2012). Paolo Legrenzi and Carlo Umiltà, tr. Frances Anderson, Neuromania: On the Limits of Brain Science. Reviewed by. *Philosophy in Review*, *32*(6), 498–500.

Logan, G. D. (2002). An instance theory of attention and memory. *Psychological Review*, *109*(2), 376.

Lupyan, G., Rakison, D. H., & McClelland, J. L. (2007). Language is not just for talking: Redundant labels facilitate learning of novel categories. *Psychological Science*, *18*(12),1077–1083 https://doi.org/10.1111/j.14679280.2007.02028.x

Lupyan, G., & Thompson-schill, S. L. (2012). *The Evocative Power of Words : Activation of Concepts by Verbal and Nonverbal Means*. *141*(1), 170–186. https://doi.org/10.1037/a0024904

Lupyan, G., & Ward, E. J. (2013). *Language can boost otherwise unseen objects into visual awareness*. https://doi.org/10.1073/pnas.1303312110/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1303312110

Manning, C., Scerif, G., & Norcia, A. (2018). *Electrophysiological correlates of motion processing and decision-making in children and adults*.

Maxfield, J. T., Stalder, W. D., & Zelinsky, G. J. (2014). Effects of target typicality on categorical search. *Journal of Vision*, *14*(12), 1.

Maxfield, J. T., & Zelinsky, G. J. (2012). Searching through the hierarchy: How level of target categorization affects visual search. *Visual Cognition*, *20*(10), 1153– 1163.

Mento, G., Scerif, G., Granziol, U., Franzoi, M., & Lanfranchi, S. (2019). Dissociating top-down and bottom-up temporal attention in Down syndrome : A neurocostructive perspective. *Cognitive Development*, *49*(December 2018), 81–93. https://doi.org/10.1016/j.cogdev.2018.12.004

Moores, E., Laiti, L., & Chelazzi, L. (2003). Associative knowledge controls deployment of visual selective attention. *Nature Neuroscience*, *6*(2), 182.

Myachykov, A., & Posner, M. I. (2005). Attention in Language. In *Neurobiology of Attention*. https://doi.org/10.1016/B978-0-12-375731-9.50057-4

Nako, R., Wu, R., Smith J., T. J., & Eimer, M. (2014). Item and category-based attentional control during search for real-world objects: Can you find the pants among the pans? *Journal of Experimental Psychology: Human Perception and Performance*, *40*(4), 1283–1288. https://doi.org/10.1037/a0036885

Peng, G., Zheng, H.-Y., Gong, T., Yang, R.-X., Kong, J.-P., & Wang, W. S.-Y. (2010). The influence of language experience on categorical perception of pitch contours. *Journal of Phonetics*, *38*(4), 616–624.

Petridis, S., Rajgarhia, V., & Pantic, M. (2015). Comparison of single-model and multiple-model prediction-based audiovisual fusion. *AVSP*, 109–114.

Pomper, R., & Saffran, J. R. (2018). *Familiar Object Salience Affects Novel Word Learning*. *00*(0), 1–17. https://doi.org/10.1111/cdev.13053

Posner, M. I., Snyder, C. R., & Solso, R. (2004). Attention and cognitive control. *Cognitive Psychology: Key Readings*, *205*.

Potter, M. C., Kroll, J. F., & Yachzel, B. (1986). Pictures in sentences: Understanding without words. *Journal of …*, *115*(3), 281. Retrieved from http://psycnet.apa.org/journals/xge/115/3/281/

Price, C. J. (2012). NeuroImage A review and synthesis of the first 20 years of PET and fMRI studies of heard speech , spoken language and reading. *NeuroImage*, *62*(2), 816–847. https://doi.org/10.1016/j.neuroimage.2012.04.062

Quadrelli, E., & Turati, C. (2016). Origins and development of mirroring mechanisms: A neuroconstructivist framework. *British Journal of Developmental Psychology*, *34*(1), 6–23.

Resende, B. (2019). Infants' Characteristics and Skills: Dissolving the Nature/Nurture Dichotomy. *Trends in Psychology*, *27*(1), 99–111.

Saffran, J. R. (2002). *Constraints on Statistical Language Learning*. *196*, 172–196. https://doi.org/10.1006/jmla.2001.2839

Schmidt, J., & Zelinsky, G. J. (2009). Search guidance is proportional to the categorical specificity of a target cue. *The Quarterly Journal of Experimental Psychology*, *62*(10), 1904–1914.

Shafto, C. L., Conway, C. M., Field, S. L., & Houston, D. M. (2012). Visual sequence learning in infancy: Domain-general and domain-specific associations with language. *Infancy*, *17*(3), 247–271.

Soto, D., & Humphreys, G. W. (2007). *Automatic Guidance of Visual Attention From Verbal Working Memory*. *33*(3), 730–737. https://doi.org/10.1037/00961523.33.3.730

Thelen, E., & Smith, L. B. (1996). *A dynamic systems approach to the development of cognition and action*. MIT press.

Thomas, M. S. C., & Karmiloff-Smith, A. (2003). Modeling language acquisition in atypical phenotypes. *Psychological Review*, *110*(4), 647.

Vickery, T. J., King, L., Treisman, A., Wolfe, J., & Duncan, J. (2018). *Setting up the target template in visual search*. (2005), 81–92. https://doi.org/10.1167/5.1.8

Walenchok, S. C., Hout, M. C., & Goldinger, S. D. (2016). Implicit object naming in visual search : Evidence from phonological competition. *Attention, Perception, & Psychophysics*. https://doi.org/10.3758/s13414-016-1184-6

Waxman, S. R., & Gelman, S. A. (2009). *Early word-learning entails reference , not merely associations*. (May), 258–263. https://doi.org/10.1016/j.tics.2009.03.006

Wolfe, J. M., & Horowitz, T. S. (2004). *What attributes guide the deployment of visual attention and how do they do it? 5*(June), 1–7.

Wolfe, J. M., Horowitz, T. S., Kenner, N., Hyle, M., & Vasan, N. (2004a). *How fast can you change your mind ? The speed of top-down guidance in visual search*. *44*, 1411–1426. https://doi.org/10.1016/j.visres.2003.11.024

Wolfe, J. M., Horowitz, T. S., Kenner, N., Hyle, M., & Vasan, N. (2004b). How fast can you change your mind? The speed of top-down guidance in visual search. *Vision Research*, *44*(12), 1411–1426. https://doi.org/10.1016/j.visres.2003.11.024

Xu, F., & Tenenbaum, J. B. (2007). Word learning as Bayesian inference. *Psychological Review*, *114*(2), 245–272. https://doi.org/10.1037/0033295X.114.2.245

2. LANGUAGE AS A COGNITIVE TOOL

Language acquisition is a complex process supported by brain maturation and by experience individuals make in their environment. Since the prenatal period, the auditory system is sufficiently developed to allow linguistic sounds to be perceived by the fetus (Moon, Lagercrantz, & Kuhl, 2013). Therefore, the linguistic sounds are familiar acoustic stimuli for newborns, who will gradually tune to a specific linguistic pattern depending on the social environment they are immersed in. Infants progressively become selective and accurate in attending and discriminating the phonemes of their native language (Maurer & Werker, 2014). They must be able to segment the continuous flow of speech to detect regularities in sound patterns of their mother tongue (phonemes) hence, to crack words embedded in the on-going speech (Kuhl, 2004; Mattys & Jusczyk, 2001; Saffran, Aslin, & Newport, 1996). The speech segmentation and the detection of regularities cannot take place in the absence of a fine attentional system early biased towards auditory stimuli (Robinson & Sloutsky, 2004). The auditory dominance allows efficient processing of auditory stimuli that have a shorter duration and rapidly overlap in time, compared to visual stimuli that can furnish a stabile and more extended presentation by facilitating visual processing. The auditory dominance allows infants to allocate higher cognitive resources towards the auditory channel (compared to other perceptual channels) by contributing to lexical acquisition. Evidence obtained in studies focusing on the first year of life indicate that infants of 6 months of age already exhibit lexical knowledge measured in terms of vocabulary size, that progressively "explodes" around 9-to-12 months of age (Bergelson & Swingley, 2012; Fenson et al., 1994; Swingley & Aslin, 2000).

A fundamental step in the trajectory of lexical acquisition is the computation of the referential link between language and real entities. Some authors suggested that analogical comparison processes are needed for acquiring concepts like the referential link (Christie & Gentner, 2010; Doumas & Hummel, 2013). Learning the reference that binds words and real entities is challenging and shows asymmetry in how some words are acquired earlier than others. In fact, infants learn names of objects and living beings before verbs and prepositions (Gentner, 1982; Gentner & Boroditsky, 2001; See Figure 5).

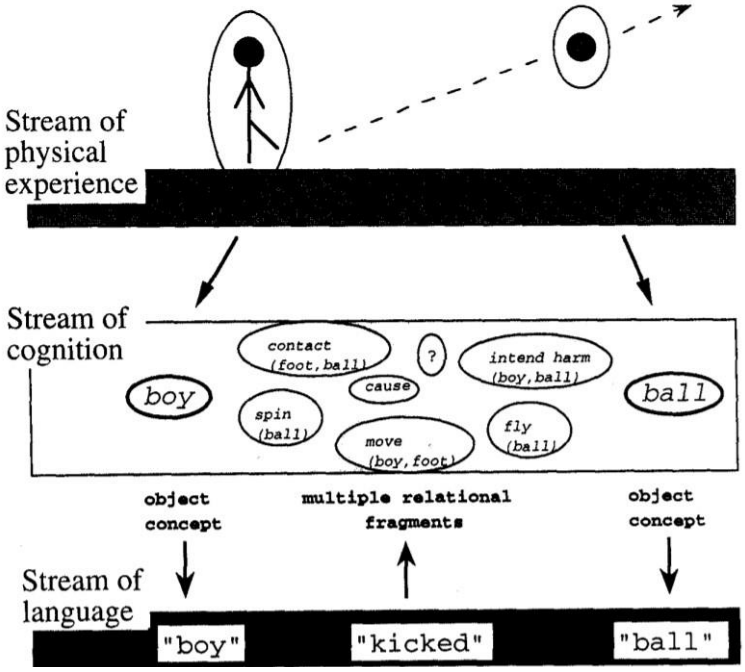


Figure 5. Interaction of language and experience (Gentner & Boroditsky, 2001).

Why do infants learn nouns first? Despite the frequent exposure infants have with certain types of words, for example, verbs related to their own motor experience, they learn object names more easily (Waxman & Guasti, 2009). The word-object pair dominance is a constant in language acquisition across many languages, which shows a clear tendency to master and learn new names in the first place rather than new adjectives or verbs, and this suggests an evolutionary priority of the lexical class (Waxman, Senghas, & Benveniste, 1997). It can, therefore, be affirmed that this nominal asymmetry of the very first vocabulary of the infant, is strictly dependent on perception (Caselli, Casadio, & Bates, 1999): the names of living beings and objects in most cases have a concrete and direct reference. Another factor that could explain the nominal asymmetry lies in its conceptual simplicity: compared to other linguistic elements through the support provided by stable pictorial information, referents are less ambiguously represented (Gentner, 1982).

The Dominance's model (Gentner, 1982) seems to fit this perceptual-attentive-linguistic dynamics. It has been introduced to explain the nominal asymmetry by considering the class to which words belong to. In traditional linguistics, open-class refers to words which group can be enlarged by adding new valid members (nouns, adjectives), whereas closed-class refers to those small number of functional words whose role is mainly to organise meaningful words in sentences (e.g., prepositions). Interestingly, closed-class words show higher frequency compared to open-class words. Thus, if the statistical co-occurrence and/or auditory dominance thoroughly explains word segmentation and the process of reference, then infants should show a closed-class superiority effect. However, what it is observed is a parametric open-class superiority that can be better explained by perceptual representations. Gentner and Boroditsky (2001) proposed this interactive and dynamic model to define language acquisition based on a continuum of dominance between more general cognitive mechanisms and more specific linguistic functions, by including perceptual and linguistic-mediated representations needed to achieve the adult speaker's performance. This model expects the early lexical acquisition to be dependent on domain-general mechanisms that gradually specialise during development. Note that this definition matches that of domain-relevant mechanisms proposed by the Neuroconstructivist approach (see Chapter 1 and Figure 6).

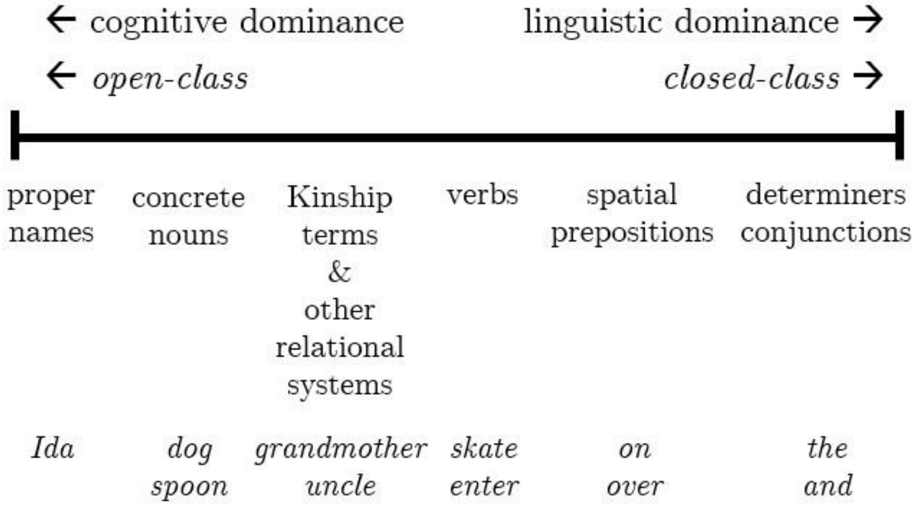


Figure 6. The Division of Dominance (Gentner, 1982).

An essential step for lexical understanding is, therefore, the learning of the association that bounds a word to its visual referent (Bavin et al., 2008; Waxman et al., 1997). Word-object pairs acquisition is a fundamental precursor of linguistic development; a more detailed explanation will be given below.

2.1. Word-Object pairs

What role does language play during cognitive development? Many studies have shown that the comprehension of the referential link between a word and a specific referent does influence the mental representation of real objects in infancy. For example, some studies have shown that labels, unlike other acoustic stimuli, influence visual perception (for a review, see Robinson, Best, Deng, & Sloutsky, 2012), visual categorisation (Althaus & Westermann, 2016; Balaban & Waxman, 1997) and mental representation of visual objects (Twomey & Westermann, 2018). In the vast literature on the subject, it is possible to distinguish at least two main hypotheses about the role played by linguistic representations on perceptual processes.

On the one side, the unsupervised feature-driven hypothesis (Robinson et al., 2012) considers linguistic sounds similar to visual features such as the colour or the shape of an object. This approach expects a visual object to be rapidly processed in the absence of any paired sound compared to an arbitrary sound-object presentation. Indeed, the arbitrary sound adds complexity to the visual object by slowing down visual processing. Thus, such an approach interprets any effect of linguistic sounds to be explained by means of bottom-up mechanisms (Sloutsky & Fisher, 2004; Sloutsky & Lo, 1999). Robinson and Sloutsky (2007) tested 14-month-olds in a continuous familiarisation with audio and video stimuli and compared the time of processing (i.e., looking time) familiar (which remained constant over the trials) and novel objects (which changed in each trial), presented simultaneously to familiar and novel sounds. To verify the role played by the linguistic sounds intended only as auditory stimuli, the two objects (familiar and new) could be presented without any sound, accompanied by a non-linguistic stimulus (sound of a laser) or by a linguistic stimulus (word). The results showed longer processing time spent looking at the familiar objects in the condition where infants were presented with familiar sounds, independently of the linguistic information. A further series of experiments focused on the influence of linguistic (Sloutsky & Robinson, 2008) and non-linguistic auditory information (Robinson & Sloutsky, 2004) on the processing of visual information (fixation times) using a familiarization-test paradigm in 10- and 16-month-old infants (see Figure 7.). The results revealed that in 10-month-old infants, concomitant novel linguistic sounds interfered with visual processing of familiar visual stimuli (compared to the processing of novel visual stimuli paired with familiar linguistic sounds). However, by pre-familiarizing 16-month-olds with non-linguistic sounds, any interference effect vanished and the findings showed a comparable performance to that obtained by 10-month-old infants with linguistic sounds. The authors concluded that, during visual processing, part of the interference effect triggered by linguistic and non-linguistic sounds depends on similar and more general processes of auditory processing and is mediated by the degree by which sounds are familiar to participants. According to this explanation, any facilitatory effect of linguistic compared to non-linguistic condition does not depend on a specific role of labels; instead, it relies on a general familiarity effect lead by linguistic sounds that co-occur more and more frequently along development. The unsupervised feature-driven approach presented above fits into a domain-general perspective that finds in auditory perception and sound-object association a univocal explanation of any word facilitatory effect, in infancy.

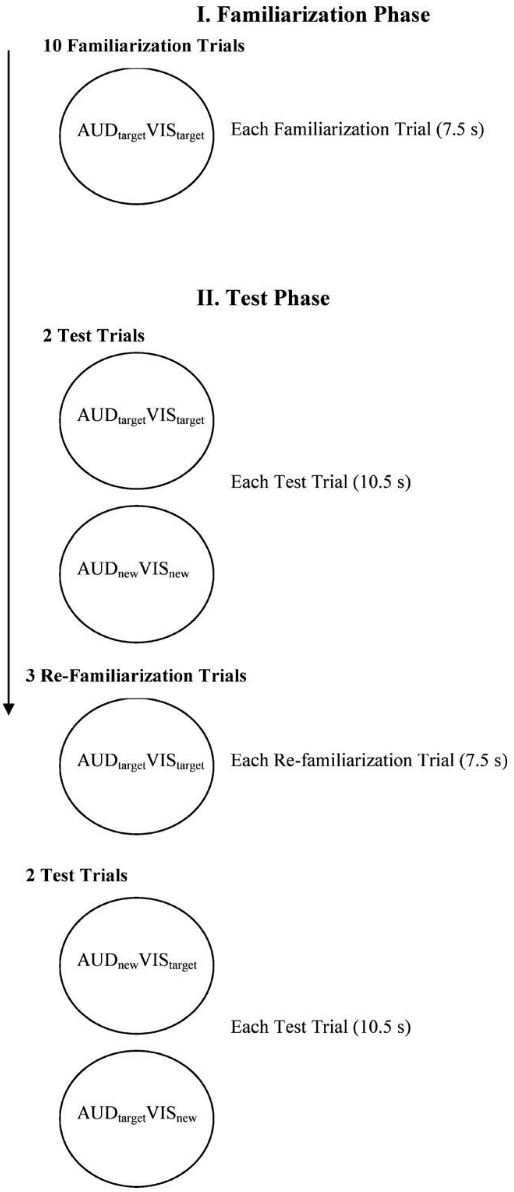


Figure 7. Procedure by Sloutsky and Robinson (2008).

On the opposite side, the supervised name-based hypothesis (Plunkett, 2011) proposes that linguistic labels preferentially guide visual tasks in a supervised top-down manner. This approach opposites a simple associative explanation of lexical acquisition by suggesting that the referential link prompt by labels does shape visual representation itself in a way that linguistic representations become an integral part of learning (Lupyan & Lewis, 2017; Perry & Samuelson, 2011; Waxman & Gelman, 2009). Therefore, in contrast with a pure associative explanation and in line with a semantic bias, this approach expects words to point to symbolic meaning (categories), whereas it expects non-linguistic sounds to point to other perceptual features they are associated with (i.e. co-occurrence with the perceptive entity of experience), independently from the stimulus familiarity. This scenario is explained by the flexibility and generativity offered by the linguistic system, compared to the auditory system (Waxman & Gelman, 2009). Balaban and Waxman (1997) were among the first that investigated in a controlled experiment, the generative potential prompted by words during a categorisation task in 9-month-old infants. In this study, the infants were pre-familiarized with images paired with either words or tones. Fixation times were analysed in the familiarisation phase and during the test phase, in which an item of the familiarised category was presented simultaneously with a new stimulus belonging to the same category. The results showed that the word condition brought more infants to look at the element of the new category for longer times than the element of the familiar category, compared to the tone condition. Fixation times from the familiarisation phase showed that the two groups of participants (word and tone conditions) did not exhibit any difference. This evidence led the authors to conclude that a general attention capture could not explain the observed effect on the part of auditory stimuli, but rather by a facilitatory effect of linguistic representation supervising the categorisation process. Similar results were reported by Fulkerson and Waxman (2007) with infants of 6- and 12- months of age tested in a categorisation task in a familiarization-test paradigm. Again, the task involved two familiarisation conditions: with tones and with words. The task elicited an effect of preference towards the novel exemplar at both ages, but only in the condition in which the images were paired with the word and not when they were paired with sinewave transformed tones. Overall the authors' interpretations are in line with the idea that in the early stages of cognitive development, linguistic sounds do exert an active and peculiar role in visual categorisation processes thus, the contribution of words during infancy should be placed at a more conceptual level rather than being considered among other co-occurring auditory features of objects. The claim of the second perspective presented above does not deny the presence of associative mechanisms that contribute to the acquisition of sound-object pair, but it refused to reduce linguistic influence to such mechanisms. The linguistic stimulus is supposed to directly point to the conceptual (symbolic) representation of the referent in a peculiar way (Twomey & Westermann, 2018), and any symbolic cues should trigger similar effects independently of the perceptual modality and depending on context-dependent attentional mechanisms. The referential and not the merely associative link, between the spoken word and real entities, allows, for example, to explain, infant's ability to refers to objects that are not contingently present or indicated by the context (Gleitman, 2005).

The two dominant approaches presented in this chapter, propose two opposite models to explain which mechanism promote lexical acquisition in the first years of life. Nevertheless, both approaches expect the influence of auditory sounds to lead to different outcomes along development by suggesting that no-linear dynamics rule language acquisition and attention to objects. To operatively investigate this common assumption, Chapter 4.1 investigates the influence of linguistic, pure auditory and visual representations on visual attention mechanisms through a traditional attention task (Blaga & Colombo, 2006), tailored for both adult and infant participants. This was made to analyse data coming from the develop*ing* and the develop*ed* brain during a similar task. According to the LFH (REF, see Chapter 1.2), the mental representation of the referent should be activated more effectively through linguistic sounds than through non-linguistics sounds in the adult cognitive system, by leading to a facilitatory effect on visual processing.

Finally, the inconsistent results found by comparing two approaches admit both that (a) the familiarity confound can explain any word facilitatory effect and that, nevertheless (b) the word influences visual processing in a peculiar way whereas other associated features can not. Chapter Section 4.1. and 4.2 aimed to disentangle these two open questions by investigating the role of visual and linguistic mediated representations on visual attention mechanisms, in infants and adults.

2.2. The sentence level

During communication, information structure can be used to stress the saliency of a piece of information, allowing an adequate amount of attention to be allocated to the most critical information. A few studies investigated sentence comprehension in infancy. However, it has been show that 24-months-olds infants make use of prosodic sentence cues during speech segmentation (Shady & Gerken, 1999), can generalise the rules of transitive verbs to new verbs (Gertner, Fisher, & Eisengart, 2006) and Event-Related Potential (ERP) studies found a late positivity associated to syntactic violation like subject-verb agreement (the lion/the lion in the roars) (Oberecker & Friederici, 2006). Thus, it seems that infants are already sensitive to sentence structure and may they make use of language and of the combinatory power of sentence, to organise the multisensorial inputs coming from the environment (Lewkowicz, 2014; Lewkowicz, Schmuckler, & Mangalindan, 2018; Neil, Chee-Ruiter, Scheier, Lewkowicz, & Shimojo, 2006; Perry & Samuelson, 2011).

However, adults make a more extensive use of sentences (this thesis is just a miserable example) by fully exploiting the generativity potential of natural language. Nevertheless, a few groups of studies have investigated perceptual representations triggered by sentences. The main paradigms that have been used to investigate linguistic-mediated representation are the Visual World Paradigm (see Chapter 1.4) and the Sentence Picture Verification task (see

Chapter 4.4). By means of an early version of the SPV, Potter Valian & Faulconer (1977) asked participants to compute the pragmatic implications of a sentence and to verify a picture or a word match. The time required to verify the sentence in the two conditions suggested that even though modality-specific representations should be computed in such a task (pictorial vs linguistic representation triggered by probes) no difference emerged by comparing word and pictorial stimuli verification times (Potter, Valian, & Faulconer, 1977). This null result opened the way to further investigations on pictorial representation triggered by linguistic cues. Further and more recent SPV studies with ERPs have been focused mainly on the latency and amplitude of the ERPs components reflecting semantic integration (i.e. N400) to investigate sentence pragmatic computation depending on sentences structure like sentence polarity (Tian, Breheny, & Ferguson, 2010); see Chapter 4.4). However, also visuospatial ERPs components may offer useful functional interpretation to investigate the computation of the pragmatic meaning triggered by sentence cues.

Different information structure can result in different priority map of the target template by overcoming the modular boundary between visual and language functions. It seems the case of adjectives which order influences the meaning computation of the whole sentence, according to traditional theories of semantics (for a detailed discussion see Cinque, 2010) as well as to computational-semantics (Baroni & Zamparelli, 2010). For example, adjectives in nominal position (e.g. *The **big** cat is black*, i.e. that directly point to the subject) and adjective in predicative position (e.g. *The big cat is **black***, i.e. that are specified after the verb) are supposed to differently affect the priority map upon which the sentence's meaning (and its mental representations) are computed. The influence of adjective order triggered by sentence cues can be investigated thanks to the SPV to disentangle the mechanisms affecting the verification by looking at the early stage of the time course of attention deployment after the picture onset.

Moreover, by utilising visuospatial components, it became possible to investigate how the sentence structure and the visual attributes that is, linguistic mediated and pictorial representations do affect attention deployment by re-prioritizing the saliency map of the attribute represented. Research on sentence comprehension has shown longer processing times and higher error rates for negative compared to affirmative sentences verification and concurrently, negative false sentences have shown to pay a smaller cost compared to negative true sentence verification times (Truth and Polarity Interaction, TPI) (Arita et al., 2012; 20Kaup, Lüdtke, & Zwaan, 2006; Scappini, Vespignani, & Delfitto, 2015; Tian et al., 2010; Tian, Ferguson, & Breheny, 2016). The TPI consistently found across experiments had led to several theories about negative template and negation representation that will be presented in section 4.4. There, it will be reported two original experiments that took advantage of search times and ERPs measures to investigate picture selection and sentence verification, depending on the sentence structure defined mainly along two acceptations: adjective order and polarity (negative/affirmative sentences). Across two experiments, the TPI was challenged as a function of the alternatives to compute negative sentence meaning. Moreover, the manipulation of the adjective order was expected to guide attention towards the named features (shape and colour) depending on the adjective position in the sentence. The investigation of the TPI and the adjective position offered an interesting insight into the question of how sentences and specifically, negative sentences trigger perceptual representation and the cascade effect on sentence-guided search.

2.1. Bibliografy

Althaus, N., & Westermann, G. (2016). Labels constructively shape object categories in 10-month-old infants. *Journal of Experimental Child Psychology*, *151*, 5–17.

Balaban, M. T., & Waxman, S. R. (1997). Do words facilitate object categorization in 9-month-old infants? *Journal of Experimental Child Psychology*, *64*(1), 3–26.

Bavin, E. L., Prior, M., Reilly, S., Bretherton, L., Williams, J., Eadie, P., … Ukoumunne, O. C. (2008). The Early Language in Victoria Study: Predicting vocabulary at age one and two years from gesture and object use. *Journal of Child Language*, *35*(3), 687–701.

Bergelson, E., & Swingley, D. (2012). At 6-9 months, human infants know the meanings of many common nouns. *PNAS*, *109*(9), 3253–3258. https://doi.org/10.1073/pnas.1113380109

Blaga, O. M., & Colombo, J. (2006). Visual processing and infant ocular latencies in the overlap paradigm. *Developmental Psychology*, *42*(6), 1069–1076. https://doi.org/10.1037/0012-1649.42.6.1069

Caselli, C., Casadio, P., & Bates, E. (1999). A comparison of the transition from first words to grammar in English and Italian. *Journal of Child Language*, *26*(1), 69– 111.

Christie, S., & Gentner, D. (2010). Where hypotheses come from: Learning new relations by structural alignment. *Journal of Cognition and Development*, *11*(3), 356–373.

Doumas, L. A. A., & Hummel, J. E. (2013). Comparison and mapping facilitate relation discovery and predication. *PloS One*, *8*(6), e63889.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., … Stiles, J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development*, *59*(5), i–185.

Fulkerson, A. L., & Waxman, S. R. (2007). Words (but not tones) facilitate object categorization: Evidence from 6-and 12-month-olds. *Cognition*, *105*(1), 218– 228.

Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. *Center for the Study of Reading. Technical Report N. 257*.

Gentner, D., & Boroditsky, L. (2001). Individuation, relativity, and early word learning. *Language Acquisition and Conceptual Development*, *3*, 215–256.

Gertner, Y., Fisher, C., & Eisengart, J. (2006). Learning words and rules: Abstract knowledge of word order in early sentence comprehension. *Psychological Science*, *17*(8), 684–691.

Kuhl, P. K. (2004). Early Language Acquisition: Cracking the Speech Code. *Nature Reviews Neuroscience*, *5*, 831–843.

Lewkowicz, D. J. (2014). *Perception of Auditory-Visual Temporal Synchrony in Human Infants Perception of Auditory-Visual Temporal Synchrony in Human Infants*. (November 1996). https://doi.org/10.1037/0096-1523.22.5.1094

Lewkowicz, D. J., Schmuckler, M. A., & Mangalindan, D. M. J. (2018). Learning of hierarchical serial patterns emerges in infancy. *Developmental Psychobiology*, *60*(3), 243–255. https://doi.org/10.1002/dev.21614

Lupyan, G., & Lewis, M. (2017). From words-as-mappings to words-as-cues : the role of language in semantic knowledge. *Language, Cognition and Neuroscience*, *0*(0), 1–19. https://doi.org/10.1080/23273798.2017.1404114

Mattys, S. L., & Jusczyk, P. W. (2001). Do Infants Segment Words or Recurring Contiguous Patterns? *Journal of Experimental Psychology: Human Perception and Performance*, *27*(3), 644–655. https://doi.org/10.1037/00961523.27.3.644

Maurer, D., & Werker, J. F. (2014). Perceptual narrowing during infancy: A comparison of language and faces. *Developmental Psychobiology*, *56*(2), 154–178. https://doi.org/10.1002/dev.21177

Moon, C. M., Lagercrantz, H., & Kuhl, P. K. (2013). Language experienced in utero affects vowel perception after birth: A two-country study. *Acta Paediatrica, International Journal of Paediatrics*, *102*(2), 156–160. https://doi.org/10.1111/apa.12098

Neil, P. A., Chee-Ruiter, C., Scheier, C., Lewkowicz, D. J., & Shimojo, S. (2006). Development of multisensory spatial integration and perception in humans. *Developmental Science*, *9*(5), 454–464. https://doi.org/10.1111/j.14677687.2006.00512.x

Oberecker, R., & Friederici, A. D. (2006). Syntactic event-related potential components in 24-month-olds' sentence comprehension. *Neuroreport*, *17*(10), 1017–1021.

Perry, L. K., & Samuelson, L. K. (2011). The shape of the vocabulary predicts the shape of the bias. *Frontiers in Psychology*, *2*, 345.

Plunkett, K. (2011). The role of auditory stimuli in infant categorization. In L. M. Oaekes, C. H. Cashon, M. Casasola, & D. H. Rakison (Eds.), *Infant perception and cognition: Recent advances, emerging theories, and future directions* (pp. 203–221). Oxford University Press, USA.

Potter, M. C., Valian, V. V, & Faulconer, B. A. (1977). Representation of a sentence and its pragmatic implications: Verbal, imagistic, or abstract? *Journal of Verbal Learning and Verbal Behavior*, *16*(1), 1–12.

Robinson, C. W., Best, C. A., Deng, W. S., & Sloutsky, V. M. (2012). The role of words in cognitive tasks: What, when, and how? *Frontiers in Psychology*, *3*, 1–8. https://doi.org/10.3389/fpsyg.2012.00095

Robinson, C. W., & Sloutsky, V. M. (2004). Auditory dominance and its change in the course of development. *Child Development*, *75*(5), 1387–1401.

Robinson, C. W., & Sloutsky, V. M. (2007). *Visual processing speed : effects of auditory input on visual processing*. *10*(6), 734–740. https://doi.org/10.1111/j.1467-7687.2007.00627.x

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-monthold infants. *Science*, *274*(5294), 1926–1928. https://doi.org/10.1126/science.274.5294.1926

Shady, M., & Gerken, L. (1999). Grammatical and caregiver cues in early sentence comprehension. *Journal of Child Language*, *26*(1), 163–175.

Sloutsky, V. M., & Fisher, A. V. (2004). Induction and categorization in young children: a similarity-based model. *Journal of Experimental Psychology: General*, *133*(2), 166.

Sloutsky, V. M., & Lo, Y.-F. (1999). How much does a shared name make things similar? Part 1. Linguistic labels and the development of similarity judgment. *Developmental Psychology*, *35*(6), 1478.

Sloutsky, V. M., & Robinson, C. W. (2008). The role of words and sounds in infants' visual processing: From overshadowing to attentional tuning. *Cognitive Science*, *32*(2), 342–365.

Swingley, D., & Aslin, R. N. (2000). Spoken word recognition and lexical representation in very young children. *Cognition*, *76*(2), 147–166.

Tian, Y., Breheny, R., & Ferguson, H. J. (2010). Why we simulate negated information: A dynamic pragmatic account. *The Quarterly Journal of Experimental Psychology*.

Twomey, K. E., & Westermann, G. (2018). Learned Labels Shape Pre-speech Infants' Object Representations. *Infancy*, *23*(1), 61–73.

Waxman, S. R., & Gelman, S. A. (2009). *Early word-learning entails reference , not merely associations*. (May), 258–263. https://doi.org/10.1016/j.tics.2009.03.006

Waxman, S. R., Senghas, A., & Benveniste, S. (1997). A cross-linguistic examination of the noun-category bias: Its existence and specificity in

French-and Spanishspeaking preschool-aged children. *Cognitive Psychology*, *32*(3), 183–218.

Waxman, S. R., & Teresa Guasti, M. (2009). Nouns, Adjectives, and the Acquisition of Meaning: New Evidence from Italian-Acquiring Children. *Language Learning and Development*, *5*, 50–68. https://doi.org/10.1080/15475440802347528

3. ORIENTING OF VISUAL ATTENTION: A WINDOW INTO MENTAL REPRESENTATION

Individuals should select regularities in a multisensory environment and map them into a referential system to acquire language (Quine, 1990; Smith & Yu, 2008; Neil, Chee-Ruiter, Scheier, Lewkowicz, & Shimojo, 2006). To select regularities means to be able to engage attention on relevant multi-sensory cues while ignoring distractors (Bahrick & Lickliter, 2000; Jusczyk, Houston, & Newsome, 1999; Kuhl, 2004; Saffran, 2002). Thus, orienting of attention mechanisms are fundamental to select meaningful items in time and space sequentially and to extract abstract rules from co-occurrent events, since infancy (Ferguson, Franconeri, & Waxman, 2018; Mitsven, Cantrell, Luck, & Oakes, 2018). Here, the focus is on the orienting of attention network and in particular on three sub-mechanisms: *engagement* that requires the attentional resources to be allocated towards a selected stimulus; *disengagement* that is the de-allocation of attentional resources from the previously selected stimulus; *shifting* that allows to move attention towards new stimuli in different directions in time and space. These mechanisms are embedded in a loop sequence that allows individuals since infancy to efficiently orient towards up-coming stimuli in the environment (Colombo, 2001; Hood & Atkinson, 1993; Posner & Cohen, 1984).

The orienting of attention in time and space allows to explore the surrounding environment and to update the reference framework in which stimuli coming from multiple sources are organised. The sensory brain areas designated to perceptual processing are not unambiguously modality-specific as shown for example, by cross-modal projections through which auditory representation do mediate visual behaviour (Von Melchner, Pallas, & Sur, 2000). Moreover, the attention system is widely distributed and show rich feed-forward connections across the whole adult brain. Petersen and Posner (2012) identify the neural network of the orienting of attention with the Posterior Attentional

System involving sub-cortical areas associated to shifting (Superior colliculus) and engagement (Pulvinar) mechanisms, and cortical areas linked to disengagement of attention (parietal lobe) and the attention shifting (temporoparietal junction the "Frontal Eye Fields", FEF) (see Figure 8).

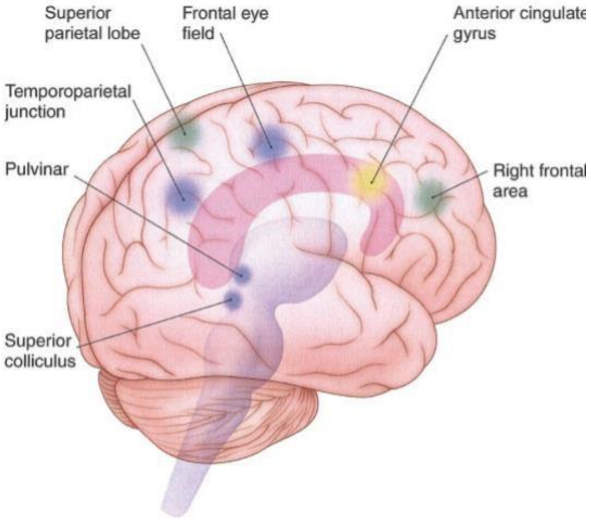


Figure 8. The Posnerian attention from Raz (2004). A sketch of the functional anatomy of the attentional networks. The pulvinar, superior colliculus, superior parietal lobe, and frontal eye fields are often activated in studies of the orienting network. The temporoparietal junction is active when a target occurs at a novel location.

Attention abilities during the first year of life has been linked to the brain maturational processes of frontal and parietal cortical areas that, in general, do require a longer developmental time compared to sensorial and sub-cortical areas (Butcher, Kalverboer, & Geuze, 2000; Matsuzawa & Shimojo, 1997). From a functional perspective, from the first to the third month of life, infants show difficulty in voluntarily shift visual attention by exhibiting prolonged visual fixations that necessarily constrain the development of further cognitive abilities. According to the obligatory attention hypothesis, the newborns' attention system is dominated by mechanisms stimulus-dependent in a bottom-up fashion (exogenous attention). Then, the oculomotor system in newborns should be mainly constrained by the salient features of stimuli such as high contrast and movement (Reynolds & Romano, 2016), which impinge the retina and then affect the Upper Colliculus response in an automatic way (Johnson, 1990). On the contrary, top-down effects on visual behaviour are expected to be very small given the insufficient support of the short- and long-term memory during information processing, as a result of the immaturity of associative and frontal areas. As a consequence, prolonged fixation and, in turn, longer latency of disengagement observed later in development can be interpreted as outcomes marking atypical brain development required by attention control and memory functioning and language acquisition (Dispaldro et al., 2013).

3.1. Disengagement of attention

The present chapter will focus on disengagement of attention as a useful and flexible mechanism which investigation can help to disentangle the effect of bottom-up and top-down mechanisms affecting the orienting of attention network in 12-months-old (Chapter 4.1) and adults (Chapter 4.2). It will be discussed that by allowing the withdrawal of attention thanks to bottom-up (exogenous attention) and controlled and top-down (endogenous attention) mechanisms, the disengagement mechanism is a suitable candidate to be a domain-relevant mechanism upon which language and visual functions rely on. Since infancy, the disengagement mechanism provides individuals with the ability to withdraw attention resources from an information unit and re-allocate them towards new upcoming stimuli in the surrounding space (Blaga & Colombo, 2006; Colombo, 2001; Posner & Cohen, 1984). This is a fundamental ability essential to voluntary shift attention in space: Without such ability, attention would remain fixed on salient stimuli, which automatically attract visual attention (Colombo, 2001; Wolfe, Horowitz, Kenner, Hyle, & Vasan, 2004).

Around four months of age, the frontoparietal areas encounter dramatic changes as a result of the competition between the early dominant processes underlying engagement of attention (obligatory looking), and those that allow to re-allocate attention in the visual field (Butcher et al., 2000). Once the processes of voluntarily re-orienting of visual attention become dominant, the frequency of disengagement decreases and stabilises, as shown by the increase of attention shifting and gaze shifting abilities from infancy to adulthood. After the first three months of life, the performance of infants improves drastically and from 6 months of age, infant's performance in voluntarily attention shift becomes qualitatively similar to those of adults (Hood & Atkinson, 1993; Matsuzawa & Shimojo, 1997).

In order to analyse the orienting of attention development, numerous studies have been carried out comparing infants, children and adults in traditional tasks sets to measure disengagement of attention from a central stimulus (Butcher et al., 2000; Cousijn, Hessels, & Kemner, 2017; Hood & Atkinson, 1993; Van der Stigchel, Hessels, van Elst, & Kemner, 2017). The main difference across ages has been mainly addressed by the difference in latency required by attention disengagement to occur, with longer times in infants than in adults (Hood & Atkinson, 1993; Van der Stigchel et al., 2017). Then, the slow developmental trajectory of disengagement of attention abilities has been related to the slow maturation of cortical areas that do include the "Frontal Eye Field" and the parietal lobe (Hood & Atkinson, 1993; Johnson, Posner, & Rothbart, 1991; Van der Stigchel et al., 2017). On one hand, the time required by information processing is influenced in a bottom-up fashion by both the individual differences in low-level processing speed (Frick, Colombo, & Saxon, 1999) and the perceptual features of stimuli (Hunnius & Geuze, 2004; Peltola, Leppänen, Palokangas, & Hietanen, 2008; Valenza et al., 2015). In fact, the longer the time is needed to process the stimuli, the higher the disengagement latencies can be, and vice versa (Frick et al., 1999).

On the other hand, information processing and, in turn, disengagement latency are influenced in a top-down fashion as shown for example, by the facilitatory effect triggered by mental representation of familiar stimuli (Blaga & Colombo, 2006). In a study, Blaga and Colombo (2006) recorded eye-movements in 3-month and 7-month-olds during an Overlap paradigm in which the presentation of a central stimulus is followed by a peripheral stimulus (see figure 2). In one condition, the central stimulus stayed the same across trials hence, it becomes more and more familiar to infants, while in the other condition, the central stimulus changed across each trial. The familiar central stimulus influenced the latencies of disengagement by facilitating orienting of attention and prompting shorter latency of disengagement compared to the condition in which the central stimulus changed across trials. In this study, the stimulus familiarity per se affected the latency of disengagement independently of the maturational level: even though 3-month-olds did show a larger effect compared to 7-month-olds infants.

The hypothesis that attentional engagement and disengagement, are influenced by the stimulus familiarity (which have been already coded and stored) suggests that the mental representation of familiar objects held in memory do modulate attention deployment in infancy. Mitsven et al. (2018) have shown that 10-month-old infants can quickly encode (in just 500 milliseconds) a visual stimulus and to use it immediately after this representation in memory to modulate orienting of attention. The authors hypothesised that like in adults, even 10month-olds can quickly create a mental representation (i.e. attentional template) in visual short-term memory (VSTM) (Oakes, Baumgartner, Barrett, Messenger, & Luck, 2013), and use this representation to guide subsequent attentional behaviour, and eye movements. The results of these study show that, during the presentation on a screen of two randomly arranged objects, 10-month-old children tend to direct their gaze more towards the not familiar object, while they pay very little attention to the object they have already seen previously. These results show that familiarity and top-down mechanisms exert an effect of orienting of attention mechanisms (engagement, disengagement and shifting) during the first year of life.

Nevertheless, the studies presented above considered exclusively visual stimuli and representation in visual short-term memory (Mitsven et al., 2018; Oakes et al., 2013) whereas no study has investigated whether the representation of multisensory stimuli (sound-object pair) in short-term memory has a similar influence on orienting of attention mechanisms and specifically on disengagement of attention, in infancy (for a similar investigation on children and adults, see Matusz, Merkley, Faure, & Scerif, 2019). Although it is widely accepted that infants around the first year of life are able to manipulate and store multisensory representations and are stable enough to modulate cognitive functioning and behaviour in even more complex tasks (e.g. categorisation tasks) (Althaus & Westermann, 2016; Capelier-mourguy, Twomey, Capelier-mourguy, Twomey, & Westermann, 2019). Few studies have investigated the effect of multisensory cues on the engagement of attention during a spatial task in children (Matusz et al., 2019). However, it is still unclear the degree at which mental representation triggered by multisensorial cues and in particular by linguistic cues, does have cascade effects on orienting of attention mechanisms. Chapter 4 will investigate, in infants and adult population, the bottom-up effect of linguistic and non-linguistic stimuli during visual stimuli encoding and their top-down effect in the Overlap paradigm, for the first time.

3.2. Looking times and pupil phasic response

The orienting behaviour has been mainly investigated through the analysis of eye movements. This methodological choice assumes that the infant's visual behaviour indirectly points to infant’s active attention deployment. Then, the analysis of oculomotor behaviour has been set to study visual attention development from birth. Among others, the Gap-Overlap (Johnson et al., 1991), is a widely used paradigm developed to investigate saccadic latency as an indirect index of disengagement of attention *per se* than other attentional mechanisms (Cousijn et al., 2017). An advantage of the Gap-Overlap paradigm is that it can be used effectively with infants as early as the first month of life (Hood & Atkinson, 1993). This paradigm involves the use of two stimuli, the first (S1) appears at the centre of the visual field and the second (S2) appears to the right or left periphery of the visual field. During the task, participants fix S1 and after the onset of S2 they shift their attention towards the latter. The paradigm includes two different conditions namely, the Gap condition in which S1 disappears as soon as S2 appears and the Overlap condition in which S1 and S2 overlap in time (see Figure 9).

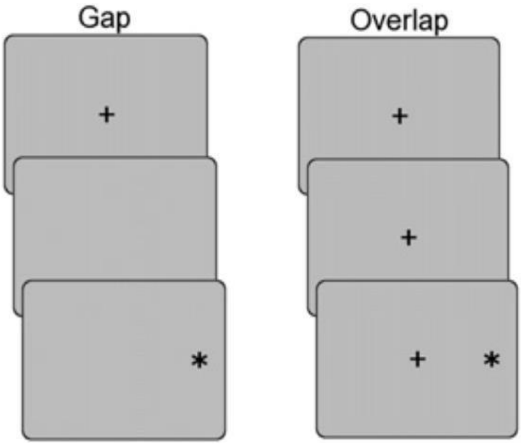


Figure 9. A gap-overlap task where a target can occur after fixation offset (gap), with the fixation remaining on-screen (overlap) from Keehen et al., 2013.

In both cases the latency times between the onset of S2 and the moment in which the participant shifts the gaze from S1 towards S2 is an index of the efficiency of disengagement of attention mechanism (Cousijn et al., 2017). In the case of the Gap condition, the participant is facilitated by the disappearance of the central stimulus, from which the disengagement occurs automatically and passively (Hood & Atkinson, 1993). Thus, in the Gap condition, attentional resources are free and ready to be reoriented as soon as S2 appears. Therefore, a pure disengagement from S1 is not necessary by leading to shorter latency times compared to the Overlap condition (Blaga & Colombo, 2006; Colombo, 2001). Whereas in the Overlap condition, disengagement of attention from S1 is required in order to attention shift to occur towards S2. Thus, the disengagement latency in this condition, compared to the previous one, is usually longer as it requires extra-time to actively disengage attention from S1 (Cousijn et al., 2017). The Overlap interference effect has been observed and replicated in numerous studies as a pure index of disengagement, and it shows to decrease during childhood (Cousijn et al., 2017; Elsabbagh et al., 2013; Matsuzawa & Shimojo, 1997) as a result of an improved efficiency in the orientation network attention and the inhibitory control system (Hood & Atkinson, 1993; Johnson et al., 1991). However, the analysis of eye-movements (looking times and saccade latency) cannot fully disentangle the interpretation of top-down and bottom-up mechanisms (endogenous vs exogenous attention) because the functional interpretation of such measures is affected at least by two main confounds: The variability of individual oculomotor speed depending on maturational processes and the fatigue and fuzziness that reduce eye-movements towards the experimental stimuli as a function of time in infant population.

On the contrary, variations in pupil diameter provide a useful indirect measure of the processing demand (Blaser, Eglington, Carter, & Kaldy, 2014; Brisson et al., 2013; Sirois & Jackson, 2012) that accurately reflects the time course of information processing, and it is not affected by fatigue over trials because only a few seconds of exposure to the stimulus is enough to detect attention fluctuations locked to a specific event (Hepach & Westermann, 2016). Although pupil dilation and constrictions mainly depend on luminance variation in the environment, it is possible to investigate psychological processes, such as attention, arousal and cognitive load by controlling for it (Beatty, 1982; Karatekin, Couperus, & Marcus, 2004; Porter, Troscianko, & Gilchrist, 2007), like emerged from numerous studies conducted with adults in the past, and more recently on infant populations (Hepach & Westermann, 2016). Changes in pupil size have physiological basis involving both the autonomic and somatic nervous systems associated with activation of the subcortical structure of the locus coeruleus, and it is considered an unbiased and involuntary marker of nervous system activity, as brain activity recorded on the scalp with EEG (see Figure 10.; Hepach & Westermann, 2016; Patwari et al., 2012) and cognitive functions such as attention, arousal, and cognitive load (Beatty, 1982; Karatekin et al., 2004; Porter et al., 2007). Adult studies report that, in addition to the amount of information held within memory (cognitive load), pupil size predicts accuracy of memory representations by distinguishing, for example, novel stimuli from familiar stimuli presentation (Laeng, Sirois, & Gredebäck, 2012).

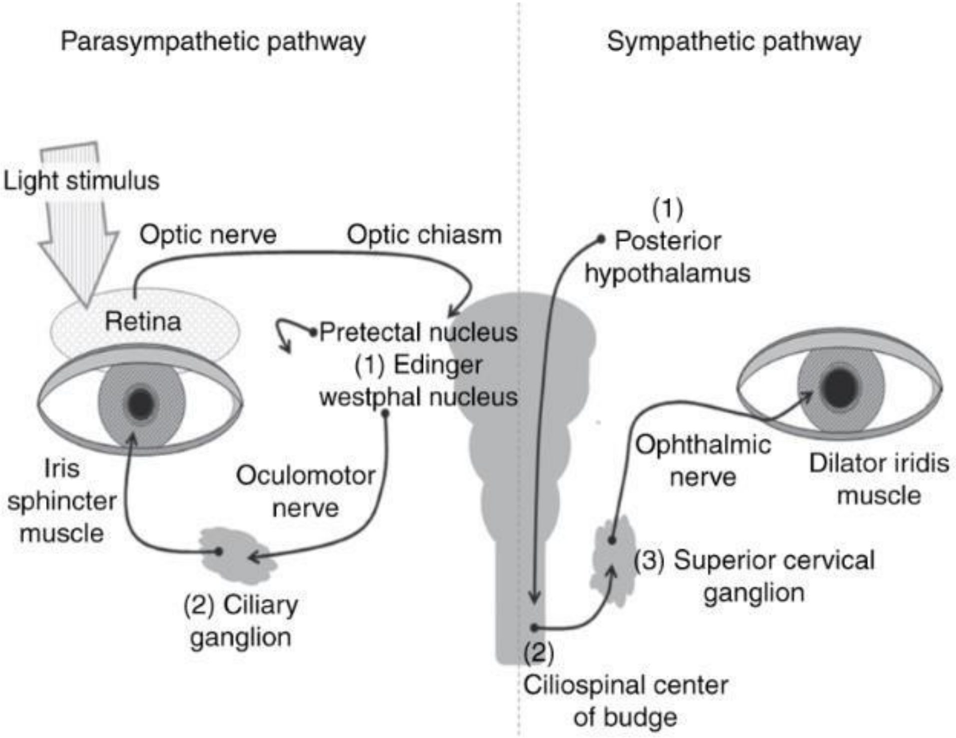


Figure 10. The autonomic system's parasympathetic and sympathetic pathways involved in pupillary control from Patwari et al., 2012.

For instance, in a coding-retrieval task, adult participants were asked to encode in memory a list of words that should be recalled in a test phase. During the test phase, variation of pupil dilation helped disentangle recalled from forgotten words after the word onset and before behavioural response (familiar/novel), by suggesting that linguistic mental representation of efficiently encoded stimuli do elicit larger pupil dilation compared to stimuli encoded less efficiently and hence, forgotten during the test (Kucewicz et al., 2018). Thus, the transient and event locked pupil phasic response reflect active engagement on perceptual events (Laeng et al., 2012) and provide a signal for 'strength of memory' (Otero, Weekes, & Hutton, 2011), and 'neural novelty' (Naber, Frässle, Rutishauser, & Einhäuser, 2013).

By using the combined measure of eye-movements and pupil dilation response in Chapter 4, it will be presented the investigation of visual encoding during linguistic and non -linguistic stimulation and the top-down effects of mental representation of encoded information (word-object and tone-object) on disengagement of attention mechanism. According to previous results, higher pupil dilation should trigger subsequent faster disengagement of attention from the familiar stimulus as a function of 'strength of memory' in adults (Otero, Weekes, & Hutton, 2011). This measure help disambiguating the role of visual, auditory and linguistic information by showing which one in a controlled experiment, triggers a more stable mental representation of familiar objects and in turn, facilitate orienting of attention.

3.3. Target selection and identification

Does disengagement of attention require identification of whatever is in focus? In the Overlap paradigm participants are not required to attend to a target to perform the task actively and this powerful methodological aspect allows to test infant's efficiency in such a specific mechanism. Nonetheless, the presentation of visual stimuli during the Overlap paradigm do trigger attention towards visual stimuli by affecting disengagement of attention from S1, both in infants and adults. However, it is not entirely clear if the engagement on S1 reflects spatial engagement or target identification. Hence, the familiarity effect on disengagement mechanisms might not depend on target identification.

Zivony & Lamy (2016) proposed an integrative metaphor of attention that reconciles this dissociation. The authors asked adult participants to perform three cueing tasks by manipulating the compatibility between the cue and the target colour, which was invariably a red-letter. Participants should report the identity of the red letter independently from the spatial cue that in half of the trials showed the same colour (compatibility effect). This manipulation allowed the authors to investigate, in the same task, the two processes of spatial engagement and target selection, separately. The RTs and accuracy findings showed that valid spatial cue efficiently captured attention on the target compared to invalid spatial cue. However, attention capture was triggered by spatial cues that did share the same colour of the target compared to that which did not share the target's colour. The authors concluded that both stimulus-driven (i.e., valid spatial cue) and goal-driven factors (i.e., red valid spatial cue) determined attentional priority: shifting attention to a location (engagement) did not necessarily entails that all features at that location were processed (selection). These findings suggest that bottom-up and top-down mechanisms exert a mutual and dynamic influence on strategies employed during visual search. Concurrently, target selection is not a necessity during the Overlap task.

On the one hand, stimulus-driven effects during visual search have been widely investigated and show that different visual features do prioritise some stimuli with respect to others during target selection (Theeuwes, 2010; Wolfe et al., 2004). On the other hand, some theories ascribed to top-down information represented in memory the primary guidance role during visual attention deployment (Folk & Remington, 2006). Somewhere in the middle, the Reverse Hierarchy Theory (RHT, Ahissar & Hochstein, 2004) try to explain spatial attention deployment and object identification mechanisms in a top-down and goal-directed fashion due to feedforward connections between subcortical and cortical areas that reflect the interface of explicit and implicit mechanisms during visual search. In this framework, top-down mechanisms operate on mental representation depending on hierarchy saliency triggered by the features of the target and by goal-directed and controlled saliency (letter identification).

3.4. The N2pc and SPCN: electrophysiological marker of template-guided search

It is well known that working memory representations of target features such as a colour and shape (Wolfe, 2007) or even, labels (Wolfe et al., 2004), guide visual attention deployment (Duncan & Humphreys, 1989). According to the biased competition account (Desimone & Duncan, 1995), a visual representation in working memory depends on sensory neurons that fire prior to search, and that lead to a competitive advantage for a stimulus that matches the previously activate representation. In addition to biasing selection towards target features, top-down mechanisms allow to deploy attention away from distractor features as in the case of negative template (Arita, Carlisle, & Woodman, 2012; see chapter 4.2). Most theories of attention propose that goal-driven search is implemented through such mental representation also called attentional template in VWM, that bias perceptual competition towards matching items (Desimone & Duncan, 1995; Hollingworth, Matsukura, & Luck, 2013). The N2pc component – a negative ERP deflection that shows its maximum amplitude on PO7/PO8 electrodes 180-200 ms after target onset in the contralateral hemifield – is an index of such a visuospatial selection and it is has been interpreted as reflecting the interaction between top-down and bottom-up mechanisms (Eimer, 1996). The N2pc amplitude increases as a function of the match between the attentional template activated before visual search and the on-line target (Eimer & Kiss, 2008). Its functional interpretation of active selection mechanism has been disambiguated from other alternatives explanation such as distractor suppression mechanism by showing that its amplitude selectively increases depending on the target presence and it is independent from the distractors homogeneity and/or spatial proximity

(Mazza, Turatto, & Caramazza, 2009). In addition, the analysis of its latency has been shown to be modulated by the physical features of the target, such as visual stimuli intensity (Brisson, Robitaille & Joelicouer, 2007).

The N2pc is be modulated by both bottom-up and top-down mechanisms and this candidate this component as a privileged gate to investigate linguistic effects during visual attention deployment. Indeed, the functional interpretation of such a measure allows to study the time course of target selection with high temporal resolution depending on the degree by which the cue allows to activate a stable and efficient attentional template in VWM. Nevertheless, a few studies investigated word cues effect during visual search (Nako, Smith, & Eimer, 2004). Furthermore, even fewer studies have investigated the effect of sentence structure (Caffarra, Pesciarelli, & Cacciari, 2013) through visual attention deployment mechanisms.

In chapter 4, in two studies, it was stressed the possibility to generate a mental representation of the target triggered by words and sentence. Chapter 4.1 describes two experiments in which participants were asked to perform a simple categorical search task, in which the N2pc was analysed during category item selection. It was expected that words of superordinate categories (e.g., stationery) to trigger top-down mechanisms during target selection by activating more conceptual and canonical representations of objects (the cue and the target share a few perceptual feature) compared to both words (e.g., pen) and pictures during a typical search task. Thus, the visuospatial N2pc component was pushed to a point. This manipulation was made to investigate the effects of linguistic mediated representation on mental representation in VWM.

A further experiment (see Chapter 4.2) investigated the ability of the attentional system to bias attention deployment far from distractors depending on negative template (Arita et al., 2012). In a Sentence Picture Verification (SPV) task, the sentence structure was manipulated (polarity and adjective position) to investigate whether linguistic flexibility and complexity influences the priority dimension map of perceptual template. The N2pc allowed to investigate early mechanisms of attention deployment towards single features depending on sentence structure (adjectives). In addition, the Sentence Picture Verification task requires participants to discriminate the target by computing the truth value of the sentence. Thus, this task was expected to elicit a further and later visuospatial and contralateral EEG component, namely, the subsequent sustained contralateral negativity (SPCN,) a negative ERP deflection that shows its maximum amplitude on PO7/PO8 electrodes 300-600 ms after target onset in the contralateral hemifield, which functional interpretation candidates, it as a pure marker of information maintenance in VWM to allow further identification of target(Mazza, Turatto, Umiltà, & Eimer, 2007). The SPCN component allow to disambiguate if the early prioritised map of template attributes changes over time as a function of the semantics (truth value) prompted by the later effects of the visual reference framework (picture), during an SPV task. By taking advantage of pure visuospatial tasks and of ERP components that reflect attention deployment in specific location of the visual field and top-down mechanisms, I aimed to shed light on the mutual effects between linguistic and pictorial information in VWM, during target selection and identification.

3.5. Bibliography

Ahissar, M., & Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. *Trends in Cognitive Sciences*, *8*(10), 457–464.

Althaus, N., & Westermann, G. (2016). Journal of Experimental Child Labels constructively shape object categories in 10-month-old infants. *Journal of Experimental Child Psychology*, *151*, 5–17. https://doi.org/10.1016/j.jecp.2015.11.013

Arita, J. T., Carlisle, N. B., & Woodman, G. F. (2012). Templates for rejection: configuring attention to ignore task-irrelevant features. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(3), 580.

Bahrick, L. E., & Lickliter, R. (2000). *Intersensory Redundancy Guides Attentional Selectivity and Perceptual Learning in Infancy*. *36*(2), 190–201.

Beatty, J. (1982). Task-Evoked Pupillary Responses, Processing Load, and the Structure of Processing Resources. *Psychological Bulletin*, *91*(2), 276–292.

Blaga, O. M., & Colombo, J. (2006). Visual processing and infant ocular latencies in the overlap paradigm. *Developmental Psychology*, *42*(6), 1069–1076. https://doi.org/10.1037/0012-1649.42.6.1069

Blaser, E., Eglington, L., Carter, A. S., & Kaldy, Z. (2014). Pupillometry reveals a mechanism for the Autism Spectrum Disorder (ASD) advantage in visual tasks. *Scientific Reports*, *4*, 1–5. https://doi.org/10.1038/srep04301

Brisson, J., Mainville, M., Mailloux, D., Beaulieu, C., Serres, J., & Sirois, S. (2013). Pupil diameter measurement errors as a function of gaze direction in corneal reflection eyetrackers. *Behavior Research Methods*, *45*(4), 1322–1331. https://doi.org/10.3758/s13428-013-0327-0

Butcher, P. R., Kalverboer, A. F., & Geuze, R. H. (2000). Infants' shifts of gaze from a central to a peripheral stimulus: A longitudinal study of development between 6 and 26 weeks. *Infant Behavior and Development*, *23*(1), 3–21.

Caffarra, S., Pesciarelli, F., & Cacciari, C. (2013). The interaction between language and visual spatial attention systems in grammatical gender processing. An N2pc study. *Cognitive Neuroscience*, *4*(3–4), 217–224.

Capelier-mourguy, A., Twomey, K., Capelier-mourguy, A., Twomey, K. E., & Westermann, G. (2019). *Modelling the interactions between auditory labels and attentional focus Modelling the interactions between auditory labels* *and attentional focus*. (June). https://doi.org/10.13140/RG.2.2.13799.29600

Colombo, J. (2001). The Development of Visual Attention in Infancy. *Annual Review of Psychology*, *52*(1), 337–367. https://doi.org/10.1146/annurev.psych.52.1.337

Cousijn, J., Hessels, R. S., & Kemner, C. (2017). *Evaluation of the Psychometric Properties of the Gap-Overlap Task in 10-Month-Old Infants*. *22*(4), 571–579. https://doi.org/10.1111/infa.12185

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*(1), 193–222.

Dispaldro, M., Leonard, L. B., Corradi, N., Ruffino, M., Bronte, T., & Facoetti, A. (2013). Visual attentional engagement deficits in children with Specific Language Impairment and their role in real-time language processing. *Cortex*, *49*(8), 2126–2139. https://doi.org/10.1016/j.cortex.2012.09.012

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, *96*(3), 433–458. https://doi.org/10.1037/0033295X.96.3.433

Eimer, M. (1996). The N2pc component as an indicator of attentional selectivity. *Electroencephalography and Clinical Neurophysiology*, *99*(3), 225–234.

Eimer, M., & Kiss, M. (2008). Involuntary attentional capture is determined by task set: Evidence from event-related brain potentials. *Journal of Cognitive Neuroscience*, *20*(8), 1423–1433.

Elsabbagh, M., Fernandes, J., Jane Webb, S., Dawson, G., Charman, T., & Johnson, M. H. (2013). Disengagement of visual attention in infancy is associated with emerging autism in toddlerhood. *Biological Psychiatry*, *74*(3), 189–194. https://doi.org/10.1016/j.biopsych.2012.11.030

Engelhardt, P. E., Ferreira, F., Patsenko, E. G., Engelhardt, P. E., Ferreira, F., Patsenko, E. G., … Ferreira, F. (2017). *Pupillometry reveals processing load during spoken language comprehension Rapid Communication Pupillometry reveals processing load during spoken language comprehension*. *0218*(February). https://doi.org/10.1080/17470210903469864

Ferguson, B., Franconeri, S. L., & Waxman, S. R. (2018). Very young infants learn abstract rules in the visual modality. *PLoS ONE*, *13*(1), 1–14. https://doi.org/10.1371/journal.pone.0190185

Folk, C. L., & Remington, R. (2006). Top-down modulation of preattentive processing: Testing the recovery account of contingent capture. *Visual Cognition*, *14*(4–8), 445–465.

Frick, J. E., Colombo, J., & Saxon, T. F. (1999). Individual and developmental differences in disengagement of fixation in early infancy. *Child Development*, *70*(3), 537–548. https://doi.org/10.1111/1467-8624.00039

Hepach, R., & Westermann, G. (2016). Pupillometry in Infancy Research. *Journal of Cognition and Development*, *17*(3), 359–377. https://doi.org/10.1080/15248372.2015.1135801

Hollingworth, A., Matsukura, M., & Luck, S. J. (2013). Visual working memory modulates rapid eye movements to simple onset targets. *Psychological Science*, *24*(5), 790–796.

Hood, B. M., & Atkinson, J. (1993). Disengaging visual attention in the infant and adult. *Infant Behavior and Development*, *16*(4), 405–422.

Hunnius, S., & Geuze, R. H. (2004). Developmental changes in visual scanning of dynamic faces and abstract stimuli in infants: A longitudinal study. *Infancy*, *6*(2), 231–255.

Johnson, M. H. (1990). Cortical maturation and the development of visual attention in early infancy. *Journal of Cognitive Neuroscience*, *2*(2), 81–95.

Johnson, M. H., Posner, M. I., & Rothbart, M. K. (1991). Components of visual orienting in early infancy: Contingency learning, anticipatory looking, and disengaging. *Journal of Cognitive Neuroscience*, *3*(4), 335–344.

Jusczyk, P. W., Houston, D. M., & Newsome, M. (1999). The beginnings of word segmentation in English-learning infants. *Cognitive Psychology*, *39*(3–4), 159– 207.

Karatekin, C., Couperus, J. W., & Marcus, D. J. (2004). Attention allocation in the dual-task paradigm as measured through behavioral and psychophysiological responses. *Psychophysiology*, *41*(2), 175–185.

Kucewicz, M. T., Dolezal, J., Kremen, V., Berry, B. M., Miller, L. R., Magee, A. L., … Worrell, G. A. (2018). Pupil size reflects successful encoding and recall of memory in humans. *Scientific Reports*, *8*(1), 4949. https://doi.org/10.1038/s41598-018-23197-6

Kuhl, P. K. (2004). Early Language Acquisition: Cracking the Speech Code. *Nature Reviews Neuroscience*, *5*, 831–843. https://doi.org/10.1038/nrn1533

Laeng, B., Sirois, S., & Gredebäck, G. (2012). Pupillometry: A window to the preconscious? *Perspectives on Psychological Science*, *7*(1), 18–27. https://doi.org/10.1177/1745691611427305

Matsuzawa, M., & Shimojo, S. (1997). Infants’ fast saccades in the gap paradigm and development of visual attention. *Infant Behavior and Development*, *20*(4), 449–455.

Matusz, P. J., Merkley, R., Faure, M., & Scerif, G. (2019). Expert attention: Attentional allocation depends on the differential development of multisensory number representations. *Cognition*, *186*, 171–177.

Mazza, V., Turatto, M., & Caramazza, A. (2009). Attention selection, distractor suppression and N2pc. *Cortex*, *45*(7), 879–890.

Mazza, V., Turatto, M., Umiltà, C., & Eimer, M. (2007). Attentional selection and identification of visual objects are reflected by distinct electrophysiological responses. *Experimental Brain Research*, *181*(3), 531–536. https://doi.org/10.1007/s00221-007-1002-4

Mitsven, S. G., Cantrell, L. M., Luck, S. J., & Oakes, L. M. (2018). Visual short-term memory guides infants ' visual attention. *Cognition*, *177*(July 2017), 189–197. https://doi.org/10.1016/j.cognition.2018.04.016

Naber, M., Frässle, S., Rutishauser, U., & Einhäuser, W. (2013). Pupil size signals novelty and predicts later retrieval success for declarative memories of natural scenes. *Journal of Vision*, *13*(2), 11.

Nako, R., Smith, T. J., & Eimer, M. (2004). *The Role of Color in Search Templates for Real-world Target Objects*. 1714–1727. https://doi.org/10.1162/jocn

Oakes, L. M., Baumgartner, H. A., Barrett, F. S., Messenger, I. M., & Luck, S. J. (2013). Developmental changes in visual short-term memory in infancy: Evidence from eye-tracking. *Frontiers in Psychology*, *4*, 697.

Otero, S. C., Weekes, B. S., & Hutton, S. B. (2011). Pupil size changes during recognition memory. *Psychophysiology*, *48*(10), 1346–1353.

Patwari, P. P., Stewart, T. M., Rand, C. M., Carroll, M. S., Kuntz, N. L., Kenny, A. S., … Weese-Mayer, D. E. (2012). Pupillometry in congenital central hypoventilation syndrome (CCHS): quantitative evidence of autonomic nervous system dysregulation. *Pediatric Research*, *71*(3), 280.

Peltola, M. J., Leppänen, J. M., Palokangas, T., & Hietanen, J. K. (2008). Fearful faces modulate looking duration and attention disengagement in 7-monthold infants. *Developmental Science*, *11*(1), 60–68.

Petersen, S. E., & Posner, M. I. (2012). The attention system of the human brain: 20 years after. *Annual Review of Neuroscience*, *35*, 73–89.

Porter, G., Troscianko, T., & Gilchrist, I. D. (2007). Effort during visual search and counting: Insights from pupillometry. *The Quarterly Journal of Experimental Psychology*, *60*(2), 211–229.

Posner, M. I., & Cohen, Y. (1984). Components of visual orienting. *Attention and Performance X: Control of Language Processes*, *32*, 531–556.

Quine, W. V. (1990). 0.(1960). *Word and Object*, 345–351.

Raz, A. (2004). Anatomy of attentional networks. *The Anatomical Record Part B: The New Anatomist: An Official Publication of the American Association of Anatomists*, *281*(1), 21–36.

Reynolds, G. D., & Romano, A. C. (2016). *The Development of Attention Systems and Working Memory in Infancy*. *10*(March), 1–12. https://doi.org/10.3389/fnsys.2016.00015

Saffran, J. R. (2002). *Constraints on Statistical Language Learning*. *196*, 172–196. https://doi.org/10.1006/jmla.2001.2839

Sirois, S., & Jackson, I. R. (2012). Pupil dilation and object permanence in infants. *Infancy*, *17*(1), 61–78. https://doi.org/10.1111/j.1532-7078.2011.00096.x

Smith, L., & Yu, C. (2008). Infants rapidly learn word-referent mappings via crosssituational statistics. *Cognition*, *106*(3), 1558–1568.

Theeuwes, J. (2010). Top–down and bottom–up control of visual selection. *Acta Psychologica*, *135*(2), 77–99.

Valenza, E., Otsuka, Y., Bulf, H., Ichikawa, H., Kanazawa, S., & Yamaguchi, M. K. (2015). Face orientation and motion differently affect the deployment of visual attention in newborns and 4-month-old infants. *PloS One*, *10*(9), e0136965.

Van der Stigchel, S., Hessels, R. S., van Elst, J. C., & Kemner, C. (2017). The disengagement of visual attention in the gap paradigm across adolescence. *Experimental Brain Research*, *235*(12), 3585–3592. https://doi.org/10.1007/s00221-017-5085-2

Von Melchner, L., Pallas, S. L., & Sur, M. (2000). Visual behaviour mediated by retinal projections directed to the auditory pathway. *Nature*, *404*(6780), 871.

Wolfe, J. M., Horowitz, T. S., Kenner, N., Hyle, M., & Vasan, N. (2004). *How fast can you change your mind ? The speed of top-down guidance in visual search*. *44*, 1411–1426. https://doi.org/10.1016/j.visres.2003.11.024

Zivony, A., & Lamy, D. (2016). Attentional capture and engagement during the attentional blink: A “camera” metaphor of attention. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(11), 1886.

4. EMPIRICAL STUDIES

The present chapter presents the recollection of four studies (seven empirical investigations) that aimed to shed light on the effect of linguistic cue on orienting of visual attention mechanisms. Section 4.1. and 4.2 reported two studies set to investigate top-down and on-line effect exerted by linguistic labels during disengagement of attention, by means of eye-movements and pupil variations measures in infants and adults. In section 4.3, it is presented the empirical study that aimed to disentangle the role of labels as preferential category cues during active search of visual stimuli, through behavioural and ERPs components during a category-based search. Finally, section 4.4. shows the attempt to bridge a psycholinguistic investigation in a controlled attentive task that investigated early attention deployment during sentence verification of pictorial stimuli, through behavioural and visuospatial ERPs components. The general aim of this multi-method approach is oriented by the fundamental question of investigating whether language-mediated representations prompt and shape perceptual representations during visuospatial tasks.

4.1. Word-object pairs facilitate disengagement of attention in 12-month-olds[1]

---

[1] 1 This study has been carried out in collaboration with Eloisa Valenza[a], Sofia Russo[a], Francesco Vespignani[a], and Simone Sulpizio[bc].
The main hypothesis and the methods have been preregistered and embargoed on Open Science Framework the 04/30/2018, ORCID 0000-0002-2913-8770. As of the date of submission of this research plan for preregistration, the data had not been collected. a Department of Developmental Psychology and Socialization, University of Padua, b Faculty of Psychology, Vita-Salute San Raffaele University, Milan,
Italyepartment of Psychological Sciences, c Centre for Neurolinguistics and Psycholinguistics, Vita-Salute San Raffaele University, Milan, Italy

Language acquisition requires young learners to parse the sound patterns of the speech stream, extract discrete units (e.g., word form), and link them to appropriate meaning (Mattys & Jusczyk, 2001; Jusczyk, 1997). These first steps into language acquisition, however, are not prompted by language information *per se*. In fact, infants start understanding words meaning via multimodal communication with adults who pair linguistic and not linguistic aspects (Kendon, 2004; McNeill,2005; Lakshmi, Gogate, Bolzani & Betancourt,2009; Bates, 2004). This multimodal communicative environment recruits multisensory perception and guides selective attention in linking new words to new objects (Gogate, WalkerAndrews, & Bahrick, 2001). Determining the meaning of a newly encountered word is effortful for the infant, because of the referential uncertainty inherent in everyday experience (Quine, 1960).; for example, it is unlikely that the onset of the word cat in the utterance 'Look, there is a cat' matches in time the appearance of the referent by orienting infant's attention towards the target (i.e., the cat). However, most theories of language acquisition assume that the co-occurrence of word and objects should take place for infants to efficiently acquire language (for a review, see Plunkett, 1997). Indeed, 12-month-old infants can detect the phonotactic structure of words (Stager & Werker, 1997; Friedrich & Friedrici, 2011; 2017) and map them to their valid visual referent through the so-called fast mapping (Smith & Yu, 2008; Curtin, 2009). This mapping allows infants to associate words to the visual objects co-occurring with and to link a word to its referent's representation (Waxman & Gelman, 2009; Zamuner, Fais & Werker, 2014; Voluoumanos & Werker, 2009). Evidence that fast mapping is implemented in terms of attention orientation toward specific space locations came from studies with mentioned but absent referents (Ganea, Shutts, Spelke, DeLoache, 2007). It has been shown, for example, that 12-month-old infants make use of word presence to orient their attention toward the location where they expect to find the mentioned object in the absence of visual reference (Saylor, 2004). This evidence suggests that in 12month-old infants, the referential link between a spoken word and a visual object may be represented (at least) in short-term memory and guides visual attention. However, whether and to what extent information encoded in short-term memory affects the mechanisms of orienting of visual attention is still unknown (Colombo, Shaddy, Richman, Maikranz, & Blaga, 2004). This means that we are underestimating the infant's ability to manipulate information in working memory during the first year of life (Colombo and Cheatham, 2006). Furthermore, even if previous studies account mainly for saccades anticipation, looking times and saccade latency as useful measures of visual attention and discrimination ability (Oakes, 2010), they do not allow to discriminate active attention engagement in looking behaviour (Colombo and Cheatham, 2006). Here, together with looking times and saccades latency, I took advantage of pupillometry to isolate moments of focused attention and quantify their effect (Cheng, Kaldy, & Blaser, 2019). Then, thanks to these complementary measures of attentional control (Astle and Scerif, 2011; Karatekin, 2007) in the present study I investigated focused and sustained attention towards objects during word-object mapping (Hollich, Golinkoff & Hirsh-Pasek,2007). Through controlled training, the representation represented in the memory-attention system in 12-month-old infants was manipulated to investigate whether just few exposures to a word-object pair do constrain orienting of attention mechanisms. Specifically, the hypothesis here expected words to prompt the online encoding of visual objects (Sloutsky & Robinson, 2008) by facilitating the subsequent disengagement of attention from it.

Several authors have investigated the mechanisms underlying the process of word-object mapping in infants (e.g., Stager & Werker, 1997; Ferry, Hespos & Waxman, 2010; Fennell, 2011,2006; Robinson &

Sloutsky, 2007b; Sloutsky & Robinson, 2008). Among these, only one recent study (Twomey and& Westermann, 2018) addressed the question whether pairing the object with a word affects the mental representation of the object. Twomey and Westermann (2018) asked parents of 10-month-old infants to play with them, at home over a week, with two 3D objects. One of the two objects was always presented with the same word (i.e., labelled condition) while the other object was presented alone (i.e., without a word, unlabelled condition). After the week-long training, the authors exposed the infants to a silent familiarisation phase in which the two trained objects were shown on display; a preferential looking task followed this phase. The results showed that during the silent familiarisation phase infants looked longer to the trained object in the labelled condition than to the one that was trained in the unlabeled condition. Twomey and Westermann (2018) interpreted the longer looking times as a novelty response and concluded that the word-object pair shaped the object representation. That is, the absence of a label slowed down the recognition of the visual object only when the object was previously paired with a label. Nevertheless, the interpretation of the novelty response made by the authors stays unclear because it should imply that the encoded label did exert an interference effect during further recognition of the same object presented silently. To disentangle these findings, the main aim of the present study was to investigate the cascade effects of linguistic-mediated representation during attention deployment towards familiar stimuli. With this aim looking times and pupil dilation were recorded as indirect measures of visual engagement and resource allocation, during a familiarisation training with novel Word-Object pairs and novel Object Only stimuli. (Cheng, Kaldy, & Blaser, 2019; Posner, Rothbart, Sheese, & Voelker, 2012).

Top-down information and disengagement of attention in infancy

Recently, Mitsven, Cantrell, Luck and Oakes (2018) showed that information stored in visual short-term memory guides 10-month-old infants' visual attention toward new stimuli when the preexposed stimulus and a new one overlap in time. Following these results, the present study aimed to investigate whether pre-exposed words encoded in short-term memory affected the orienting of attention from visual objects. My hypothesis is that if the previous exposure to a word-object pair shapes the representation of the object in short term memory (Twoney & Westermann, 2018), then the infant's ability to disengage his/her visual attention from the object will depend on the consistency between the information encoded in short-term memory and the information processed online. (Mitsven et al.,; 2018). Specifically, here consistency/inconsistency refers to the conditions of the Overlap task with word presence/absence (after a word-object pair familiarisation) and colour presence/absence (after a coloured object familiarisation). The scenario in which the Inconsistent condition similarly affects disengagement of attention similarly independently from the Training would suggest that words act as features of the object (Robinson & Sloutsky, 2007b; Sloutsky & Robinson, 2008; Wolfe & Horowitz, 2004). In contrast, if the Word-Object pair and the Object Only differently affect disengagement of attention, then, the two representations in the memory-attention system should different attribute priority that differently affect the early attention mechanisms, in a top-down fashion. Thus, if only the word or only the colour absence (e.g., for colour, see, Bartels & Zeki, 2000; Brouwer & Heeger, 2009; Conway, Moeller, & Tsao, 2007; Hadjikhani, Liu, Dale, Cavanagh, & Tootell, 1998; for shape, see, Malach et al., 1995; Grill-Spector, Kushnir, Edelman, Itzchak, & Malach, 1998; Kourtzi & Kanwisher, 2000, 2001) affects attention disengagement mechanisms then, it would be possible to ad interim conclude that, compared to the visual feature of the object (Wolfe & Horowitz, 2004), the word contributes from a more conceptual level to visual recognition (Waxman & Gelman, 2009; Twoney & Westerman, 2018).

The present study investigated representation abilities in 12month-old infants, an age at which infants are able to map words onto objects in a referential fashion (Saylor, 2004; Smith & Yu, 2008; Curtin, 2009, Woodward, Markman, Fitzsimmons, & Colleen, 1994) and they voluntarily orient visual attention over discrete stimuli in the visual field (Blaga & Colombo, 2006). Thus, this time window represents a good test for our hypotheses and the skills required by our test do not overload the abilities showed by infants of this age (Kenward et al., 2017). The orienting of attention network is responsible for information selection and processing of specific features of a sensory input (Rueda, Fan, McCandliss, Halparin, Gruber, Lercari, & Posner, 2004; Colombo, 2001; Posner & Cohen, 1984). Usually, attentional disengagement is investigated using an overlap paradigm, which consists in the presentation of a first stimulus (S1) followed by a second peripheral stimulus (S2). Generally, when S2 appears, the viewer interrupts the fixation on S1 (disengagement), makes a saccade from S1 to S2 (shifting) and starts a fixation on S2 (engagement). Although by the age of 4 months infants can deploy their attention easily and rapidly from the central to the peripheral stimulus (Hunnius & Geuze, 2004; Hood & Atkinson, 1993; Butcher, Kalverboer, & Geuze, 2000), disengagement latency (i.e., the time interval between the appearance of S2 and the beginning of the saccadic movement) is affected by S1 (Blaga & Colombo, 2006). Critically, disengagement is considered as the stage of spatial orienting when the processing of S1 has to be terminated before shifting attention to a new location (Posner & Petersen, 1990). Thus, we expected the consistency between the information in short term memory and S1 to fasten the processing and the disengagement of attention from S1 compared to when S1 is inconsistent with respect the representation stored in short term memory. The experiment comprised two blocks with two phases each: a training phase and a following overlap task (see Figure 1). In one block of the training phase (Word-Object), a visual object was systematically paired with a word to form S1, so that the infant map the former into the latter (Friedrich & Friederici, 2011). By contrast, in another block, infants were presented with an object as S1only (Object Only, i.e., without word pairing).

In both blocks, the overlap task (Frick, Colombo, & Saxon, 1999; Blaga & Colombo, 2006) immediately followed the training phase, with each infant being exposed to two conditions, namely a Consistent and an Inconsistent condition (see Figure 1). In the Consistent condition, the central stimulus was the same of the training phase (Word-Object pair or Object Only). In contrast, in the Inconsistent condition, the visual object was presented without one feature, i.e., silently or without colour in the first and the second block, respectively.

If the orienting of attention mechanisms are guided by the information encoded during the training, we expected to observe opposite patterns of disengagement depending on the consistency of the central stimulus (S1) presented in the overlap task (see Figure 1). Specifically, it was expected to find longer latencies of disengagement from the central stimulus, longer total looking times and higher pupil phasic response (i.e., attentional engagement, Hepach & Westermann, 2016) in the Inconsistent condition. In other words, longer looking times are expected when the presented object does not match the representation encoded in short-term memory (Lupyan, 2008; Twomey & Westermann, 2018). Instead, higher pupil phasic response was interpreted as an index of stimulus encoding during the two trainings (Cheng, Kaldy, & Blaser, 2019) and as an index of 'novelty response' (Tàmasi, McKean, Gafos, and Hohle; 2019) by comparing the two overlap task' conditions. The training comparison would give an insight into the differences in terms of mental representation built upon a word-object pair vs an object alone. Secondly, these results would demonstrate that infants rely on the kind of information encoded in short-term memory to orient their attention toward visual stimuli.

4.1.1. Methods

Participants

The pre-registered sample size planned to enrol sixty 12month-old infants (mixed for gender) born at full term, in good health, with no sensorial or neurological disorders or any familial language disorder. Infants were recruited from a database of parents of the Department of Developmental Psychology and Socialization, University of Padua; parents were contacted via mail and telephone.

Infants were exposed to both a training phase and the overlap task in two different blocks, whose order was counterbalanced among participants. One block involved the Word-Object pair followed by the overlap task. The other block involved the Object Only training followed by the overlap task (see Figure 1). The sample size was established following the rule-of-thumb of N =10 for each parameter of the multiplicative model (Austin & Steyerberg, 2015), which in a within-subject design allows to have enough statistical power to detect significant effects if any (see Oakes, 2017).

So far, we collected thirty-four 12-month-olds (SD: .84, 15 girls). Although we planned a within-subject design, not all participants achieved to complete both blocks (Word-Object e Object Only). Hence, further inclusion criteria have been used and 10 participants have been excluded because they showed, in at least one block:

- less than 7 valid trials (i.e., looking times > 100ms at the Area of Interest) during the training,

- less than 2 valid trials (valid saccadic latencies > 100ms and < 1s) per condition in the overlap task.

Thus, twenty-four 12-month-olds (M:12, SD: .7, 13 girls) were included in the sample that was composed of: - 11 infants completing both blocks (Word-Object and Object Only), (M = 11.9 months, SD = .9, 5 girls);

- 9 infants completing only Word-Object (M = 12 months, SD = .8, 6 girls);
- 4 completing only Object Only (M = 12 months, SD = .5, 2 girls). The ethics committee has approved the entire research protocol of the Department of Developmental Psychology and Socialization, University of Padua for approval (protocol number: 2423); the research was conducted in accordance to the principles elucidated in the Declaration of Helsinki. Parents provided written informed consent.

Stimuli

Visual Stimuli. The visual objects used in both the training phase and the overlap task were selected from the Novel Objects Unusual Noun (NOUN) database (Horst & Hout, 2016), which is often used in infancy research. For each object, NOUN provides measures of familiarity (i.e., the percentage of adults that reported to have already seen the object), name-ability (i.e., the percentage of adults who named the object with the same name) and colour saliency (i.e., the percentage of adults who spontaneously referred to the objects' colour (s) when asked to name the object)[2] . We used 2 objects that are expected to be unfamiliar to our participants (mean familiarity score = 16%; mean name-

---

[2] Note that frequency of color qualifiers and object novelty are correlated, with higher values of novelty associated with higher values of color saliency.

ability score = 23%, mean color saliency = 45%). All stimuli were equated for luminance (by using LightRoom software and GIMP2) to avoid any luminance confounding. Visual objects were presented 720×720 px and centred on the screen. Stimuli (and measures) are listed in Appendix.

Acoustic stimuli. The linguistic sounds were composed of 2 disyllabic pseudo-words that have been selected from the NOUN database: /coba/, and /toma/. These pseudo-words are phonotactically legal and have the most common syllabic structure in the infants's native language (i.e., the Consonant Vowel (CV) sequence). The stimuli were recorded with the Audacity software (Audacity Team, 2017; equipment: microphone SHURE PG58 and sound card M-AUDIO Fast Track). The audio stimuli were recorded by a female speaker. All stimuli were pronounced with stress falling on the penultimate syllable (which is the dominant pattern in the infants's native language; Spinelli, Sulpizio, & Burani, 2017).

Experimental Design and Procedure

Figure 11 shows the experimental paradigm.

This study has a 2 (Training: Word-Object pair vs Object Only) x 2 (Overlap task: Consistent vs Inconsistent) within-participants design (See Figure 1). Infants' eye movements and pupil dilation were recorded as response variables during both the training phase and the following overlap task.

Calibration phase: The experiment started by welcoming the parents and the infant at the lab so that they could familiarise themselves with the environment. Then, the participant sat in an infant high chair placed 60 cm from the monitor. Parents sat behind the infant's seat. At this point, the calibration phase started: five markers were presented one by one in different locations on the screen (top-left, top-right, centre, bottom-left and bottom-right).

Training phase. This phase started with the onset of a visual static object. Participants saw a different object in each block and visual objects were counterbalanced among participants and blocks. In the Word-Object block, one visual object was paired with a word. As soon as the participant reached 100 ms of fixation at the object, the auditory stimulus started automatically. In the Object Only block, one visual object was presented without any auditory stimulation. The training phase consisted of 9 trials (1 second each).

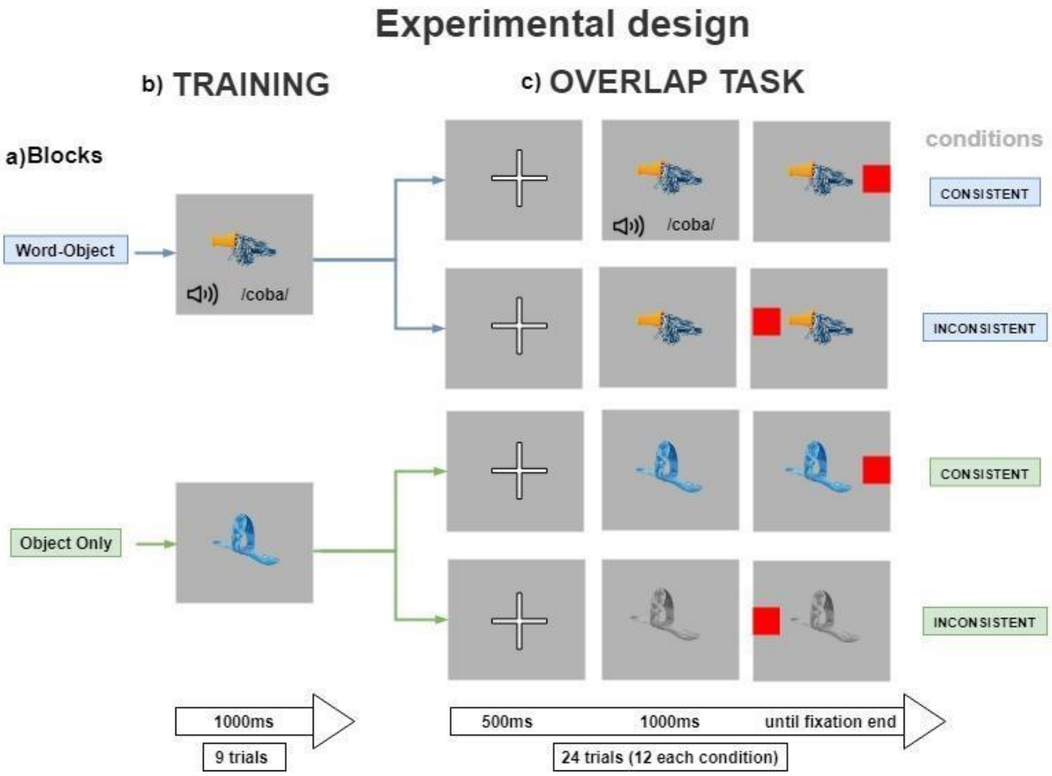


Figure 11. Participants were exposed to a) two blocks, namely the Word-Object (WO) and the Object Only (OO) block in a counterbalanced order. Both blocks started with (b) a training phase in which an object was present paired with a word or silently, respectively. Right after the training phase, (c) an overlap task started, and after a fixation cross lasting for 500 ms, the previously trained object was presented for 1000ms. Each overlap task comprised two conditions, namely a Consistent (where the object was identical to the one previously trained) and an Inconsistent in which the object was presented without a feature either the word in Word-Object or the colour in Object Only.

Overlap task. This task consisted of a total of 24 trials (12 Consistent and 12 Inconsistent) presented in a randomised order immediately after the training phase (Cousjin et al., 2018). In the Consistent condition, the object was identical to the training phase (in both the Word-Object and the Object Only blocks). In the Inconsistent condition, the object was deprived of a feature: That is, in the Word-Object block the object was presented silently (without the word), whereas, in the Object Only block, the object was presented with no colours (shade of grey).

In both conditions, if the infant was looking at the fixation cross, lasting 50ms, then, the visual object appeared for 1000ms followed by the peripheral stimulus (S2, i.e., a red square). The S2 was randomly presented either on the left or the right of the central one (5.9° distance in visual degrees from the centre to the peripheral stimuli). The trial ended after the appearance of S2, as soon as the infants moved their gaze away from the central stimulus.

The overlap task was considered valid if the infant completed at least two trials per condition. Saccadic latencies were considered valid when beginning within 0.1 and 1.0 seconds from the onset of S2 (Kenward et al., 2017).

Apparatus

The visual stimuli were presented with Open Sesame software 3.1 (Psychology Software Tools, Pittsburgh, PA) on a 27-inch monitor (Philips 300x300). A remote, infrared eye-tracking camera (Tobii X2-60 Eye-Tracker) placed directly below the screen recorded the participant's eye movements using bright pupil technology at a sampling frequency of 60 Hz. The audio stimuli were presented with two speakers (KRK rokit rp 5) placed on the right and left of the screen. The experimental session took place in a room with semi-darkness constant luminance guaranteed by a lamp positioned one meter away behind the participant. The room presented a dark curtain that isolated the participant area from the experimenter area.

Statistical Analyses

Data from the training phase and the overlap task were analysed with the R software (R Core Team, 2014). Outliers were evaluated by the Influence analysis for Generalized Mixed-Effects models (Nieuwenhuis, Grotenhuis & Pelzer, 2012). See Appendix for more details.

Looking times. For each trial of the training, it was measured the time infants spent in the AOI. Note that, each trial started as soon as participants reached 300ms of looking at the AOI, i.e. visual object, and it continued for 1 second, even if participants looked away. Then, only actual fixations in AOI were included in the analysis.

Disengagement latency. For each trial of the Overlap task the time interval between the appearance of the peripheral stimulus (S2) and the beginning of the saccadic movement towards it (i.e., fixation end). Those saccades occurring after 100ms and before 1 second were considered valid and included in the analysis (Kenward et al., 2017).

Pupillometry. I assessed differences between phases and conditions in terms of resources allocation by expecting a positive relation between increased attention and pupil diameter (Laeng, Sirois, & Gredebäck 2012; Sirois & Brisson, 2014). Pupil dilation was tracked during both the training and the overlap task. It was interpreted as an index of focused attention required for stimuli encoding in training, and for allocation of attentional resources before disengagement of attention, in the overlap task. Note that pupil dilation does not show a decrease of response over trials because it requires to look at a stimulus just for a few seconds (Jackson & Sirois, 2009; see also Sirois & Jackson, 2012 for a detailed discussion). Thus, any change of pupil dilation response during the experiment accounts exclusively for the experimental manipulation. Pupil dilation variability consists of a tonic state and a phasic response (Hepach & Westermann, 2016), and only the latter was evaluated in both the training and the overlap task. To obtain a measure of the phasic response, it was followed the Hepach and Westermann (2016) procedure: the median of raw pupil dilation values from the two eyes was calculated (if both were present); subtractive baseline correction was done by determining the median pupil size during the initial 100-ms epoch before audio onset and then subtracting that value from each data point up to 1 second (Mathôt, Fabius, Van Heusden &Van der Stigchel, 2018).

Both disengagement latency and pupil dilation data were modelled with generalized linear models (GzLMs) that allow to account for both random and fixed effects. GzLMs are an extension of the GLMs that allow to specify the distribution family. When needed, this allows us to overcome the assumptions made by GLMs that require residuals to be normally distributed, and their variability is uniform across the levels of the predictors (Fox, 2008). Since when dealing with non-negative behavioural data (e.g., reaction times or disengagement latencies) residuals are often positively skewed and heteroscedastic, these models have to be preferred with respect to classical GLM. The distribution family to use was chosen depending on an analysis of the shape of residual distributions.
All models were fitted with the glm2 package (Marschner, 2011) in R software (Team, R. C., 2013). To find the best approximation to the true model, we followed a model comparison approach, using likelihood ratio test (LTR), AIC (Akaike Information Criterion; Akaike, 1974) and AIC weight as indexes of goodness of fit. The former tests the hypothesis of no differences between the likelihoods of two nested models. The AIC and AIC weight give information on the models' relative evidence (i.e., likelihood and parsimony) so that the model with the lowest AIC and the highest AIC weight is to be preferred (Wagenmakers & Farrell, 2004). I started from the simplest model and proceeded by adding predictors (Pitt, Myung, & Zhang, 2002).

For the Training, for each dependent variable – i.e., looking times and pupil dilation – two different models were tested: the null model including the random effect of participants (dependent variable ~ (1|Participant) + residuals) and, the second model with the factor Training (Word-Object vs Object Only) as a fixed factor (dependent variable ~ Training + (1|Participants) + (1|Trial Order) + residuals). For the Overlap task, for each dependent variable – i.e., disengagement latencies and pupil dilation – four different models were tested. I started with the null model including the random effect of participants (dependent variable ~ (1|Participant) + residuals); in the second model, it was introduced the factor Training (Word-Object vs Object Only) as a fixed factor (dependent variable ~ Training + (1|Participants) + residuals); in the third model, it was introduced the second fixed factor, i.e., overlap condition (Consistent vs Inconsistent condition; dependent variable ~ Training + Overlap task condition + (1|Participants) + residuals); the fourth model also included the interaction between the two fixed factors (dependent variable ~ Training * Overlap task condition + (1|Participants) + residuals).

4.1.2. Results

Looking times

During the training phase, participants were exposed to either a Word-Object (WO) pair and/or an Object Only (OO). I analysed the time participants spent looking at the AOI defined as the visual area of the central object (720x720 pixel) during each trial. The model with the factor Training (fitted with the Gamma function, see Appendix for distribution parameter estimation, Gamma fit and influential analysis, see also Table 1) showed that participants spent more time looking at the object during the Object Only (M = 4.7., SD = 1.4) training than during the Word-Object (M = 6.9 seconds, SD = 1.3) training as shown in Figure 12.

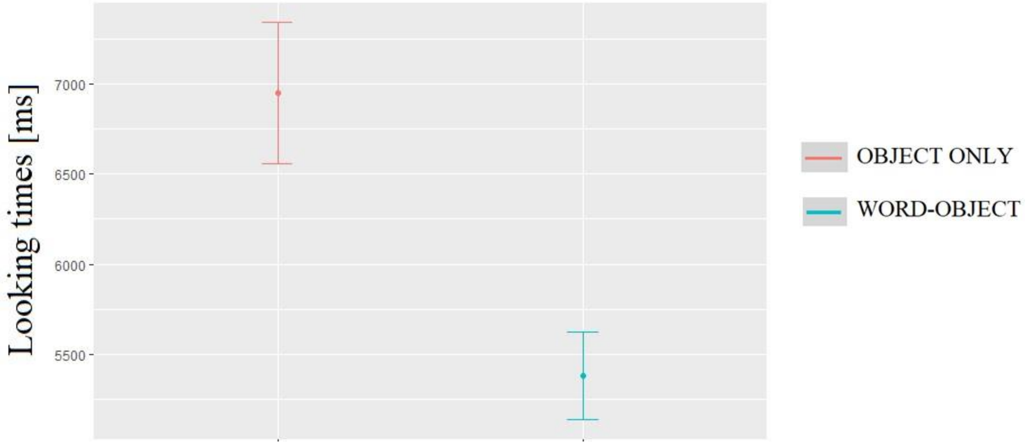


Figure 12. Total looking times in milliseconds during the Object Only and the Word-Object training; the plot reports the means and the 95% Confidence Interval (CI) for the two trainings.

The model selection in Table 1 shows that Training was a good predictor of the data. The main effect of Training showed that the Object – Only training was looked longer time compared to the Word – Object training (β = -31, SE = 0.07, t = -4.66).

models	RD	dAIC	AICw	df	χ	p	η2
Looking times ~ (1|Participants)	60	16.08	0	\	\	\	\
Looking times ~ Training + (1 | Participants)	56	0	1	1	18.85	<0.001	0.04

Table 1. Model comparison for predicting Looking times in the Training phase. RD = residual deviance, AIC = Aikake information criterion, dAIC = difference between a model's AIC and those of the best model, AICw = AIC weight, df = degrees of freedom of the chi-squared statistic, η2 = eta squared as the ratio between the chi-squared and the residual deviance of the null model.

Pupillometry

Figure 13 shows pupil dilation baseline-corrected descriptives during the two trainings on the left panel and the conditions of the Overlap tasks in the right panel. Firstly, it should be noted a decrease of focused attention after a peak around 200ms in both Trainings during an averaged trial.

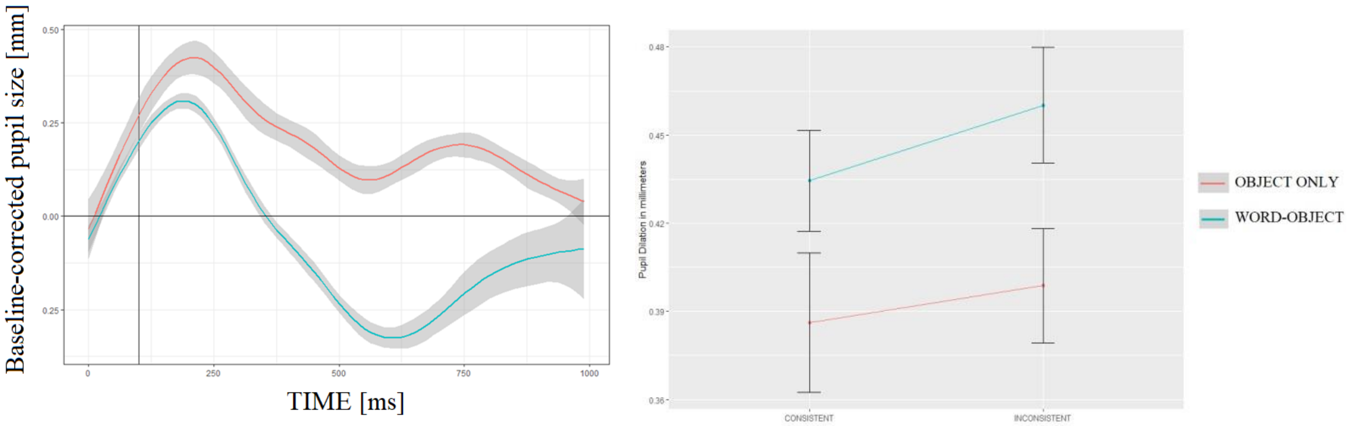


Figure 13. Baseline corrected pupil size in millimetres across time during the Training (on the left) and the following Overlap task's conditions Consistent and Inconsistent (on the right). The mean effect and the 95% Confidence Intervals (CI) are plotted.

The model selection in Table 2, showed the model with the fixed factors Training + Overlap task's condition (fitted with Gamma function, see Appendix) to better approximated our data.

models	RD	dAIC	AICw	df	χ	p	η2
Pupil dilation ~ (1|Participants)	8686	10.7	0	\	\	\	\
Pupil dilation ~ Training + (1 | Participants)	8683	7	0.02	1	5.7	0.01	0.0042
Pupil dilation ~ Overlap task's conditions+ (1 | Participants)	8674	4.9	0.05	0	2.07	<0.001	0.0015
Pupil dilation ~ Training + Overlap task's conditions + (1 | Participants)	8671	0	0.54	1	6.93	0.008	0.0051
Pupil dilation ~ Training * Overlap task's conditions + (1 | Participants)	8670	0.6	0.40	1	1.39	0.23	\

Table 2. Model comparison for predicting Saccadic latency in the Overlap tasks, RD = residual deviance, AIC = Aikake information criterion, dAIC = difference between a model's AIC and those of the best model, AICw = AIC weight, df = degrees of freedom of the chi-squared statistic. η2 = eta squared as the ratio between the chi-squared and the residual deviance of the null model.

The selected model shows a main effect of Training with infants allocating more cognitive resources on the AOI during the Object Only than the Word-Object training (β = -0.27, SE: = 0.03, t = -6.94). The main effect of Conditions conditions showed a significantly higher peak of infants' focused attention on S1 in the Inconsistent compared to the Consistent condition (β = 0.7, SE = 0.02, t = 3.5). These results suggest that, independently of the training, presenting S1 without a feature required higher resources to disengage attention compared to the Consistent condition.

Saccadic latency

The model comparison of GzLMs with Gamma family and identity link function and subjects as random effects showed no significant differences in disengagement of attention neither among Word-Object and Object Only nor between the Overlap task's conditions (see Figure 14). Nevertheless, it should be noted that, given the complexity of the statistical models, our sample size (24 participants, 370 total trial) is still too small for regression analysis to better estimate the equation of the expected effects for saccadic latency. So far, caution is needed before the planned sample (i.e., 60 participants) is reached.

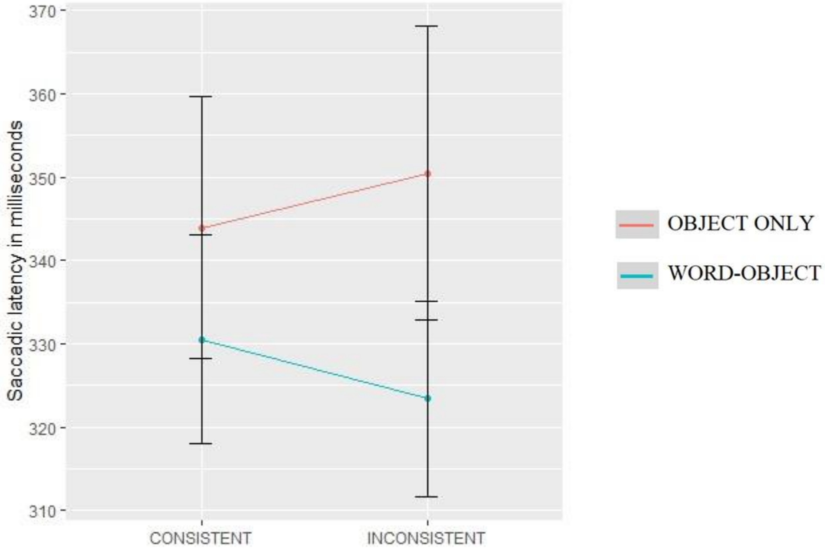


Figure 14. Interaction plot of Saccadic latency during the Overlap task's conditions further split for Object Only in red and Word-Object in green. Mean values and 95% CI are plotted.

PVB Questionnaire

In order to assess linguistic abilities of the infants, it was collected a measure of vocabulary comprehension and production by asking parents to fill the short version of the 'Primo Vocabolario del Bambino -PVB' (Caselli, Pasqualetti & Stefanini, 2007), which is the Italian version of the MacArthur Communicative Development Inventory (CDI;) (Fenson et al., 2000; Rose, Feldman & Jankowsky, 2009). This tool encompasses a multifactorial approach in investigating the early stages of language acquisition (Karmiloff-Smith,2006; Rowe and Goldin-Meadow, 2009) by measuring gestures and words comprehension ability of 8-to-24-month-olds. Critically, it has been shown that Lexical comprehension measured with the CDI positively correlates with visual memory and representational ability in infants (Rose, Feldman, & Jankowsky, 2009), thus it is a valuable measure in the present study. Lastly, it has been widely used in cross-linguistic research (for a review see Law & Roy, 2008) and validated in Italian on a sample of 583 infants from 8 to 24 months.

The sample showed an average Lexical Comprehension (LCA) age of 12 (SD = 3.3. quotient = 103%, quotient SD = 30), corresponding to the 50° percentile of the12-month-olds group of the Italian sample on which the instrument has been validated[3]. Figure 15 shows exploratory analysis that showed the higher LCA index obtained in the PVB questionnaire predicted longer time the time infants spent on the AOI during the Training phase and longer latency of disengagement of attention from S1 (i.e., saccadic latency) in the Consistent condition of the Word-Object block.

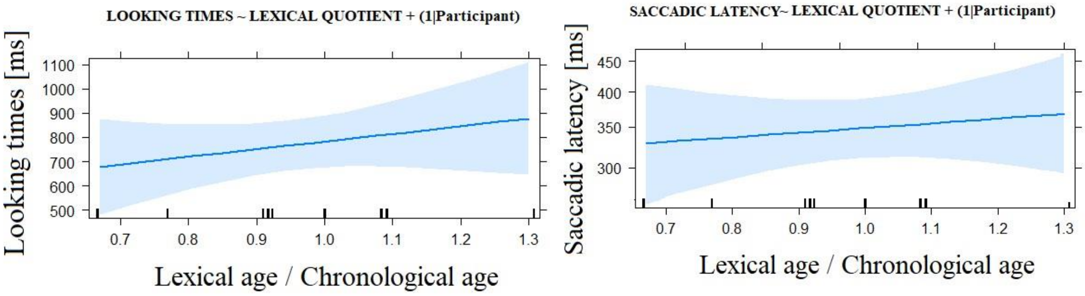


Figure 15. Effect plot of GLMM models predicting looking times in ms during the Word-Object training (on the left panel) and saccadic latency during the Consistent condition in the Word-Object block (on the right pane) for the Lexical Comprehension quotient (Lexical age/Chronological age) of the PVB questionnaire.

---

[3] For both LCA and GA, the PVB allows to compute a quotient of lexical comprehension ability as the ratio between LCA and the chronological age.

4.1.3. Discussion

To my knowledge, this study is the first attempt to fill the gap in the literature of top-down attention deployment and linguistic-mediated representation, in infancy. By capitalising on a classical paradigm – i.e., the Overlap task – twenty-four 12-month old infants were trained with object only or word-object pairs to test whether the presence of labels affected the encoding of novel stimuli and the subsequent disengagement of attention from familiar stimuli. By collecting multiple eye-tracking measures – i.e., looking times, saccade latency and pupil dilation –, it was possible to analyse complementary indirect measures of implicit and controlled attentional engagement and disengagement (Gredebäck, Johnson, & von Hofsten, 2009).

An essential premise to the discussion of the present results is that, in order to have reliable parameter estimations, I planned to test sixty infants. So far, we collected thirty-four infants; thus, even if our sample size already achieves the median in the related field (Oakes, 2017), these results should be considered as preliminary.

The preliminary findings coming from the Training showed that infants spent more time looking at the object AOI during the Object Only compared to the Word-Object training. Moreover, pupil dilation analysis confirmed a higher level of sustained attention in the former compared to the latter. These findings suggest that the word may act as a useful feature during visual information encoding (Robinson & Sloutsky, 2007b; Sloutsky & Robinson, 2008) by requiring less time and less resources allocation on the object. Further analysis considering the Lexical comprehension quotient (LCq) detected two different trends. Infants with a broader vocabulary showed shorter looking times in the Word-Object pair and longer looking times in the Object Only training than to infants with a lower score on the same scale. These results are in accordance to evidence showing that infants resulting in smaller vocabulary size need more time to encode a word-object linkage compared to infants with a higher vocabulary size (Rose et al, 2009).

On the other hand, infants with a better short-term memory can hold more information by showing an advantage in manipulating the auditory information into meaningful units (Rose et al.; 2009). A further clue to the facilitation lead by the word-object block can be obtained by observing the percentage of drop-out of participants: most infants were able to complete the Word-Object block (20 children out of 24, 83 %), while a few group managed to complete the Object Only block (15 children out of 24, 62%). This data can be considered as a further indication that the during infancy labels and therefore the multimodal presentation of word-Object is more interesting and facilitates the encoding. Further investigation of the vocabulary size predicting performance of orienting of attention may gain insight into the interactive mechanisms that operate on active learning (Perry & Samuelson, 2011).

Results coming from the Overlap task are still weak in terms of power. Disengagement of attention latency has not shown any difference between training and overlap task conditions. However, the vocabulary size emerged to be a good candidate predictor of saccadic latency in the Consistent condition of the word-object block. The ability to show a large vocabulary size might be interpreted as reflecting more rapid encoding and greater facility at disengaging attention (Colombo, 1993; Colombo, Mitchell, Coldren, & Freeseman, 1991; Freeseman, Colombo, & Coldren, 1993; Frick, Colombo, & Saxon, 1999; Jacobson, Jacobson, Sokol, Martier, & Ager, 1993).

As a complementary measure of disengagement of attention, pupil dilation analysis showed higher focused attention deployment following the Word Object compared to the Object Only, and during the Inconsistent then the Consistent condition. Greater focused attention after the Word-Object training gives insights into difference in terms of mental representations built upon multisensory and visual-only information, by showing that the Word-Object pair training made infants engaged more on the object, compared to the Object Only training even after the training (Robinson & Sloutsky, 2004). Finally, it should be notice that any difference in the Consistent condition after the Object Only compared to the Inconsistent condition after the Word-Object training (physically identical) show that the kind of information infants were previously trained with differently affected subsequent attention deployment in a top-down fashion (Mitsven, Cantrell, Luck & Oakes, 2018). Any difference between these two conditions relies uniquely on the information infants were previously exposed to and their representational ability (Rose et al., 2009). To conclude, the results of the Training suggested that the word-object pair required less resources allocation to be encoded compared to the object only. In the Overlap task, saccadic latency differences were predicted only by LCq, whereas pupil dilation analysis detected differences both between trainings and between the Overlap task's conditions by showing that the mental representation of a word-object triggered subsequent attention deployment towards the object more than a mental representation built on visual information only. It is expected that the future findings of the final sample (N = 60) will corroborate a more conceptual role of word in the representation of objects in short-term memory and subsequent orienting of attention, compared to visual features of objects (i.e. colour).

One limitation of this study relies on the fact that the lexical comprehension ability was measured with the LC quotient of the PVB questionnaire that presents the typical limitations of instrument of this nature (Fenson et al., 2000). Moreover, this experiment is not able to disentangle whether the results suggest a general advantage driven by multisensory information (Gogate, Walker-Andrews, & Bahrick, 2001; Flom, & Bahrick, 2007) or instead they claim for a pure linguistic effect. To this aim, we planned experiment 2 that by implementing the same paradigm will compare Word-Object vs Tone-Object pairs, i.e. sine-wave tones matched to the pseudowords in amplitude and duration (Fulkerson and Waxman, 2007).

4.1.4. Bibliografy

Akaike, H. (1974). A new look at the statistical model identification. IEEE transactions on automatic control, 19(6), 716-723.

Astle, D. E., & Scerif, G. (2011). Interactions between attention and visual short-term memory (VSTM): What can be learnt from individual and developmental differences?. Neuropsychologia, 49(6), 1435-1445.

Austin, P. C., & Steyerberg, E. W. (2015). The number of subjects per variable required in linear regression analyses. Journal of clinical epidemiology, 68(6), 627636.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1–7. 2014.

Blaga, O. M., & Colombo, J. (2006). Visual processing and infant ocular latencies in the overlap paradigm. Developmental Psychology, 42(6), 1069.

Butcher, P. R., Kalverboer, A. F., & Geuze, R. H. (2000). Infants' shifts of gaze from a central to a peripheral stimulus: A longitudinal study of development between 6 and 26 weeks. Infant Behavior and Development, 23(1), 3-21.

Cheng, C., Kaldy, Z., & Blaser, E. (2019). Focused attention predicts visual working memory performance in 13-month-old infants: A pupillometric study.

Developmental cognitive neuroscience, 36, 100616.

Colombo, J. (2001). The development of visual attention in infancy. Annual review of psychology, 52(1), 337-367.

Colombo, J., & Cheatham, C. L. (2006). The emergence and basis of endogenous attention in infancy and early childhood. In Advances in child development and behavior (Vol. 34, pp. 283-322). JAI.

Colombo, J., Shaddy, D. J., Richman, W. A., Maikranz, J. M., & Blaga, O. M. (2004). The developmental course of habituation in infancy and preschool outcome. Infancy, 5(1), 1-38.

Curtin, S. (2009). Twelve-month-olds learn novel word–object pairings differing only in stress pattern. Journal of Child Language, 36(5), 1157-1165.

Fennell, C. T. (2006). Object familiarity affects 14-month-old infants' use of phonetic detail in novel words. In Poster presented at the biennial international conference on infant studies, Kyoto, Japan.

Fennell, C. T. (2012). Object familiarity enhances infants' use of phonetic detail in novel words. Infancy, 17(3), 339-353.

Fenson, L., Pethick, S., Renda, C., Cox, J. L., Dale, P. S., & Reznick, J. S. (2000). Shortform versions of the MacArthur communicative development inventories. Applied Psycholinguistics, 21(1), 95-116.

Ferry, A. L., Hespos, S. J., & Waxman, S. R. (2010). Categorization in 3-and 4-monthold infants: an advantage of words over tones. Child development, 81(2), 472-479.

Fox, J. (2015). Applied regression analysis and generalized linear models. Sage Publications.

Frick, J. E., Colombo, J., & Saxon, T. F. (1999). Individual and developmental differences in disengagement of fixation in early infancy. Child development, 70(3), 537-548.

Friedrich, M., & Friederici, A. D. (2011). Word learning in 6-month-olds: fast encoding–weak retention. Journal of Cognitive Neuroscience, 23(11), 3228-3240.

Friedrich, M., & Friederici, A. D. (2017). The origins of word learning: Brain responses of 3-month-olds indicate their rapid association of objects and words. Developmental science, 20(2), e12357.

Ganea, P. A., Shutts, K., Spelke, E. S., & DeLoache, J. S. (2007). Thinking of things unseen: Infants' use of language to update mental representations. Psychological Science, 18(8), 734-739.

Gogate, L. J., Walker-Andrews, A. S., & Bahrick, L. E. (2001). The intersensory origins of wordcomprehension: an ecological–dynamic systems view. Developmental Science, 4(1), 1-18.

Hepach, R., & Westermann, G. (2016). Pupillometry in infancy research. Journal of Cognition and Development, 17(3), 359-377.

Hollich, G., Golinkoff, R. M., & Hirsh-Pasek, K. (2007). Young children associate novel words with complex objects rather than salient parts. Developmental Psychology, 43(5), 1051.

Hood, B. M., & Atkinson, J. (1993). Disengaging visual attention in the infant and adult. Infant Behavior and Development, 16(4), 405-422.

Horst, J. S., & Hout, M. C. (2016). The Novel Object and Unusual Name (NOUN) Database: A collection of novel images for use in experimental research. Behavior research methods, 48(4), 1393-1409.

Hothorn, T., Bretz, F., Westfall, P., & Heiberger, R. (2008). Multcomp: simultaneous inference for general linear hypotheses. R Package Version 1.0-3.

Hunnius, S., & Geuze, R. H. (2004). Developmental changes in visual scanning of dynamic faces and abstract stimuli in infants: A longitudinal study. Infancy, 6(2), 231-255.

Jackson, I., & Sirois, S. (2009). Infant cognition: going full factorial with pupil dilation. Developmental science, 12(4), 670-679.

Jusczyk, P. W. (1997). Finding and remembering words: Some beginnings by English-learning infants. Current Directions in Psychological Science, 6(6), 170-174.

Karatekin, C. (2007). Eye tracking studies of normative and atypical development. Developmental review, 27(3), 283-348.

Karmiloff-Smith, A. (2006). The tortuous route from genes to behavior: A neuroconstructivist approach. Cognitive, Affective, & Behavioral Neuroscience, 6(1), 9-17.

Kenward, B., Koch, F. S., Forssman, L., Brehm, J., Tidemann, I., Sundqvist, A., ... & Gredebäck, G. (2017). Saccadic reaction times in infants and adults: Spatiotemporal factors, gender, and interlaboratory variation. Developmental psychology, 53(9), 1750.

Laeng, B., Sirois, S., & Gredebäck, G. (2012). Pupillometry: A window to the preconscious?. Perspectives on psychological science, 7(1), 18-27.

Law, J., & Roy, P. (2008). Parental report of infant language skills: A review of the development and application of the Communicative Development Inventories. Child and Adolescent Mental Health, 13(4), 198-206.

Lupyan, G. (2008). From chair to" chair": A representational shift account of object labeling effects on memory. Journal of Experimental Psychology: General, 137(2), 348.

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. Behavior research methods, 44(2), 314-324.

Mathôt, S., Fabius, J., Van Heusden, E., & Van der Stigchel, S. (2018). Safe and sensible preprocessing and baseline correction of pupil-size data. Behavior research methods, 50(1), 94-106.

Mattys, S. L., & Jusczyk, P. W. (2001). Phonotactic cues for segmentation of fluent speech by infants. Cognition, 78(2), 91-121.

Mitsven, S. G., Cantrell, L. M., Luck, S. J., & Oakes, L. M. (2018). Visual short-term memory guides infants' visual attention. Cognition, 177, 189-197.

Nelder, J. A., & Baker, R. (1972). Generalized linear models. Wiley Online Library.

Nieuwenhuis, R., te Grotenhuis, M., & Pelzer, B. (2012). Influence. ME: Tools for detecting influential data in mixed effects models. R journal 4 (2). 38–47.

Oakes, L. M. (2017). Sample size, statistical power, and false conclusions in infant looking-time research. Infancy, 22(4), 436-469.

Pitt, M. A., Myung, I. J., & Zhang, S. (2002). Toward a method of selecting among computational models of cognition. Psychological review, 109(3), 472.

Plunkett, K. (1997). Theories of early language acquisition. Trends in cognitive sciences, 1(4), 146-153.

Posner, M. I., & Cohen, Y. (1984). Components of visual orienting. Attention and performance X: Control of language processes, 32, 531-556.

Posner, M. I., Rothbart, M. K., Sheese, B. E., & Voelker, P. (2012). Control networks and neuromodulators of early development. Developmental psychology, 48(3), 827.

R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, 2014. URL http://www.Rproject.org/

Robinson, C. W., & Sloutsky, V. M. (2007). Visual processing speed: Effects of auditory input on visual processing. Developmental science, 10(6), 734-740.

Robinson, C. W., & Sloutsky, V. M. (2008). Effects of auditory input in individuation tasks. Developmental Science, 11(6), 869-881.

Rose, S. A., Feldman, J. F., & Jankowski, J. J. (2009). A cognitive approach to the development of early language. Child development, 80(1), 134-150.

Rowe, M. L., & Goldin-Meadow, S. (2009). Differences in early gesture explain SES disparities in child vocabulary size at school entry. Science, 323(5916), 951-953.

Rueda, M. R., Fan, J., McCandliss, B. D., Halparin, J. D., Gruber, D. B., Lercari, L. P., & Posner, M. I. (2004). Development of attentional networks in childhood. Neuropsychologia, 42(8), 10291040.

Saylor, M. M. (2004). Twelve-and 16-month-old infants recognize properties of mentioned absent things. Developmental Science, 7(5), 599-611.

Sirois, S., & Brisson, J. (2014). Pupillometry. Wiley Interdisciplinary Reviews: Cognitive Science, 5(6), 679-692.

Sirois, S., & Jackson, I. R. (2012). Pupil dilation and object permanence in infants. Infancy, 17(1), 61-78.

Sloutsky, V. M., & Robinson, C. W. (2008). The role of words and sounds in infants' visual processing: From overshadowing to attentional tuning. Cognitive Science, 32(2), 342-365.

Smith, L., & Yu, C. (2008). Infants rapidly learn word-referent mappings via crosssituational statistics. Cognition, 106(3), 1558-1568.

Spinelli, G., Sulpizio, S., & Burani, C. (2017). Q2Stress: A database for multiple cues to stress assignment in Italian. Behavior research methods, 49(6), 2113-2126.

Stager, C. L., & Werker, J. F. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. Nature, 388(6640), 381.

Sullivan, J. W., & Horowitz, F. D. (1983). Infant intermodal perception and maternal multimodal stimulation: Implications for language development. Advances in infancy research.

Twomey, K. E., & Westermann, G. (2018). Learned Labels Shape Pre-speech Infants' Object Representations. Infancy, 23(1), 61-73.

Vouloumanos, A., & Werker, J. F. (2009). Infants' learning of novel words in a stochastic environment. Developmental psychology, 45(6), 1611. W.V.O. QuineWord and object. MIT Press, Cambridge (1960).

Wagenmakers, E. J., & Farrell, S. (2004). AIC model selection using Akaike weights. Psychonomic bulletin & review, 11(1), 192-196.

Waxman, S. R., & Gelman, S. A. (2009). Early word-learning entails reference, not merely associations. Trends in cognitive sciences, 13(6), 258-263.

Wolfe, J. M., & Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it?. Nature reviews neuroscience, 5(6), 495.

Woodward, A. L., Markman, E. M., & Fitzsimmons, C. M. (1994). Rapid word learning in 13-and 18-month-olds. Developmental psychology, 30(4), 553.

Woodward, A., & Hoyne, K. (1999). Infants' learning about words and sounds in relation to objects. Child development, 70(1), 65-77.

Zamuner, T. S., Fais, L., & Werker, J. F. (2014). Infants track word forms in early word–object associations. Developmental science, 17(4), 481-491.

4.2. Disengagement of spatial attention as a window into word-object representation[4]

According to the Label feedback hypothesis, learning new words is useful not only for communication but also for further recognition of objects (Lupyan, 2012; Lupyan & Spivey, 2007; Lupyan & Thompson-schill, 2012; Rakison & Lupyan, 2008). In fact, in our everyday experience even during passive looking like, during a walk back home, individuals can recognise objects in the surrounding environment i.e., cars, trees, people, even in the absence of explicit goal-driven tasks. In that case, the mental representation, the recollection of some features like the colour, the shape and the verbal label of such items (James, 1890), becomes salient via top-down mechanisms thanks to feed-back/forwards connections that do integrate in a coherent scene the visible item. Nevertheless, object identification is not a mandatory step during such tasks (Zivony & Lamy, 2016) yet it can occur depending on stimulus-driven and/or goal-directed mechanisms even in the absence of any task instruction. In that case, the ability to recognise objects became "a hybrid visual-linguistic experience" (Lupyan, 2012).

Top-down mechanisms have been investigated mainly with regards of goal-directed, task-driven control (Hayhoe et al.,2003; Jovancevic et al., 2006), and passive object recognition (Neider and Zelinsky, 2006; Henderson et al., 2009; Tatler et al., 2010). In the present study, it was investigated the latter process and particularly, how top-down information interfere with passive object recognition when the on-line object differs in some diagnostic features with

---

[4] This study has been carried out in collaboration with, Simone Sulpiziobc , Francesco Vespignania, Sofia Russoa , and Eloisa Valenzaa.

a Department of Developmental Psychology and Socialization, University of Padua, b Faculty of Psychology, Vita-Salute San Raffaele University, Milan, Italyepartment of Psychological Sciences, c Centre for Neurolinguistics and Psycholinguistics, Vita-Salute San Raffaele University, Milan, Italy than its familiar representation. Here, diagnostic features refer to the colour, the shape and label of the object. Apart from the colour that is univocally considered an essential feature that guide object selection (Wolfe & Horowitz, 2004), evidence suggests that the label-shape consistency plays a fundamental role during object identification (Lupyan & Ward, 2013). A more substantial facilitatory effect of labels has been observed during on-line label presentation whereas, by adding a delay (≥ 1600 ms) between the word and the object onset the facilitatory influence of labels decreases (Lupyan & Spivey, 2007).

Nevertheless, the Label-Feedback hypothesis assumes that visual objects are processed in a memory-based fashion. That is, to get head with the object identification the memory-attentional interface tries to estimate the best fit between bottom-up information and top-down reference frameworks (Logan, 2002) and linguistic labels cue to these reference frameworks. Thus, any off-line facilitatory effect of labels should emerge by comparing objects which labels are unknown to objects which labels are known. Experiment 1 investigated this hypothesis that is consistent with evidence suggesting a facilitatory effect lead by matching features (Nako, Wu, Smith J., & Eimer, 2014), and moved a step forward. Experiment 1 aimed to collect evidence about any difference existing between linguistic and visual effect triggered by online presentation of familiar and unfamiliar stimuli, by investigating the cascade effect of label representation on early mechanisms of attention deployment: the milestone of high-level cognitive functions (Petersen & Posner, 2012). Experiment 2 was set to disambiguate if the facilitatory effect of labels on attention mechanisms rely on a general multisensory advantage (Bahrick & Lickliter, 2000) or instead it can be ascribed to a pure linguistic advantage (Edmiston & Lupyan, 2015). In two experiments, participants were asked to learn the co-occurrence of Word-Object (WO), Object Only (OO) and Tone-Object (TO) pairs by expecting the WO to facilitate disengagement of attention from the visual object, compared to the OO and TO representations.

The disengagement of attention is a mechanism of the orienting network that allows the individuals to move their attentional focus from an object or a spatial location towards a new object or spatial location (Petersen & Posner, 2012). This mechanism can be efficiently investigated by using the Overlap task (Van der Stigchel, Hessels, van Elst, Kemner, 2017), which consists in the presentation of a central stimulus (S1) followed by the onset of a peripheral stimulus (S2). After a fixation on S1, individuals are instructed to reach S2 by doing a saccade towards it as soon as possible. It has been widely shown that the introduction of a temporal gap between the offset of S1 and the onset of S2 (gap condition) decreased saccadic reaction time (saccadic latency) compared to when S1 remained visible (overlap condition) in the order of 220 msec or more (slow regular saccades). Thus, in the overlap task to the gap task, the attention shift from S1 to S2 requires extra time (saccadic latency), even when participants are instructed not to pay attention to S1. In addition, control experiments have shown that the attention shift during the Overlap task could not be explained by effects of anticipation or warning during the task, because without warning signals and under randomised conditions express saccades occur as well (for a review, see Fischer & Breitmeyer, 1987). For a further discussion on disengagement of attention mechanisms, see Chapter 3.

In the present study, after a training phase, it was implemented an Overlap task to study the effects of verbal labels on the activation of the target representation (see Lupyan & Thompson-Schill, 2012) by expecting the attentional shifting towards S2 to depend on the information consistency of S1. That is, the variability of fixation times towards S1 while S2 overlaps it in time, should indicate that the direction of gaze and the target selection mechanisms are concurrently operating in the same portion of space (Saslow, 1967; Fischer & Breitmeyer, 1987). Thus, if attention selection (object identification) take place in such a task (Zivony & Lamy, 2016) the time needed by saccades to occur should reflect the match/mismatch between the features presented online and those previously stored in memory.

In the present study, attention engagement on S1 was measured by taking advantages of two complementary measures of attention deployment, namely, saccadic latency and pupil dilation. The pupil dilation phasic response is an automatic response to cognitive load (Van Engen & McLaughlin, 2008) that allows to disentangle attention engagement and identification (Mathôt, 2018). Pupil phasic variation reflects active engagement on visual stimuli during both stimuli encoding and recall of item representation stored in memory (Kucewicz et al., 2018; Attar, Schneps & Pomplun, 2013) and it helps to disambiguate the interpretations of eye movement measures, that do not necessarily reflect active looking. Privitera, Renninger, Carney, Klein and Aguilar (2008), by using a Rapid Sequence Visual Presentation, investigated whether the simple appearance of a target triggered pupil dilation, in a visual search task. They found that the target presence triggered higher pupil dilation. Critically, the latency of the dilation onset occurred between 300 and 700 msec after target presentation, a time range fully comparable with time window found with ERPs components related to further identification of the target (e.g., Eimer, 1996; Mazza, Turatto, Umiltà & Eimer, 2007). Interestingly, studies that investigated correlations at the intraindividual level hence, less sensitive to confounds (van Steenbergen & Band, 2013) consistently suggest that pupil dilation is associated with improved behaviour in conditions that require inhibitory control by showing, for example, larger preparatory dilations associated with faster saccades (Wang et al., 2012).

Visual attention is directed to entire objects rather than to individual features (Luck & Vogel, 1997; Eimer, 2017) and spatial attention is allocated serially to one object at time, such that the updating of mental representations is an important aspect of the deallocation of attention from an object before the attentional selection of a new object occurs (Treisman & Gelade, 1980). It involves the integration between information stored in memory and the updating in response to on-line information (Miyake et al., 2000; Morris & Jones, 1990). Thus, it was expected that the more the on-line object matched the familiar representation, the shorter the latency of disengagement to be.

4.2.1. Methods

Participants

Sixty-three adults (32 female, mean age = 25, SD = 3.6,) were enrolled. The inclusion criteria for all participants were to be in good health, to have no sensorial or neurological disorders or any familial language disorder. Three subjects were excluded from the analysis because they reached less than 80% accuracy in the ORt (Lupyan & Thompson-Schill, 2012).

The sample size was fixed to sixty participants according to a rule-of-thumb (N = 10 for each parameter of the multiplicative model), thus as the within-participants design should allow to have enough statistical power to detect significant effects if any (Austin & Steyerberg, 2015).

The ethics committee has approved the entire research protocol of the Department Developmental psychology and socialisation, University of Padova (protocol number: 2423); the research was conducted in accordance to the principles elucidated in the Declaration of Helsinki. Participants provided written informed consent.

Stimuli

Visual and auditory stimuli were counterbalanced across participant and blocks.

Visual Stimuli. The visual objects used in both the training phases and the overlap task were selected from the Novel Objects Unusual Noun (NOUN) database (Horst & Hout, 2016). For each object, NOUN provides measures of familiarity (i.e., the percentage of adults that reported to have already seen the object), name-ability (i.e., the percentage of adults who named the object with the same name) and colour saliency (i.e., the percentage of adults who spontaneously referred to the objects' colour (s) when asked to name the object). Six objects expected to be unfamiliar to our participants were used (mean familiarity score: 7%; mean name-ability score: 23%, mean colour saliency: 45%)[5]. All stimuli were equated for luminance (by using LightRoom software and GIMP2) to avoid any luminance confounding during pupil variation recording. Visual objects size will be presented 720×720 px and centred on the screen.

Acoustic stimuli. The pseudowords used as objects' labels were 6 disyllables selected from the NOUN database: /coba/, /gade/, /kita/, /pabe/, /reda/, and /toma/. Pseudowords were all phonotactically legal and had both the most common syllabic structure (i.e., the Consonant-Vowel sequence) and the most common stress pattern (i.e., penultimate stress, Spinelli, Sulpizio, & Burani, 2017) of the participants' native language. For each pseudoword a sequence was created by repeating each word three times with a 500ms pause after each label e.g. */coba/ - pause - /coba/ - pause - /coba/ - pause*, so as each sequence lasted 2 seconds.

Experimental Design and Procedure

This study had a 2 (Training: Word-Object pair vs Visual Object) x 3 (Overlap task: Consistent, Inconsistent and Deprived) within-participants design (see Figure 16). The experiment consisted of two bocks with two phases each, the training and the overlap task; during both phases, participants' eye movements and pupil dilation were recorded.

---

[5] Note that frequency of color qualifiers and object novelty are correlated, with higher values of novelty associated with higher values of color saliency.

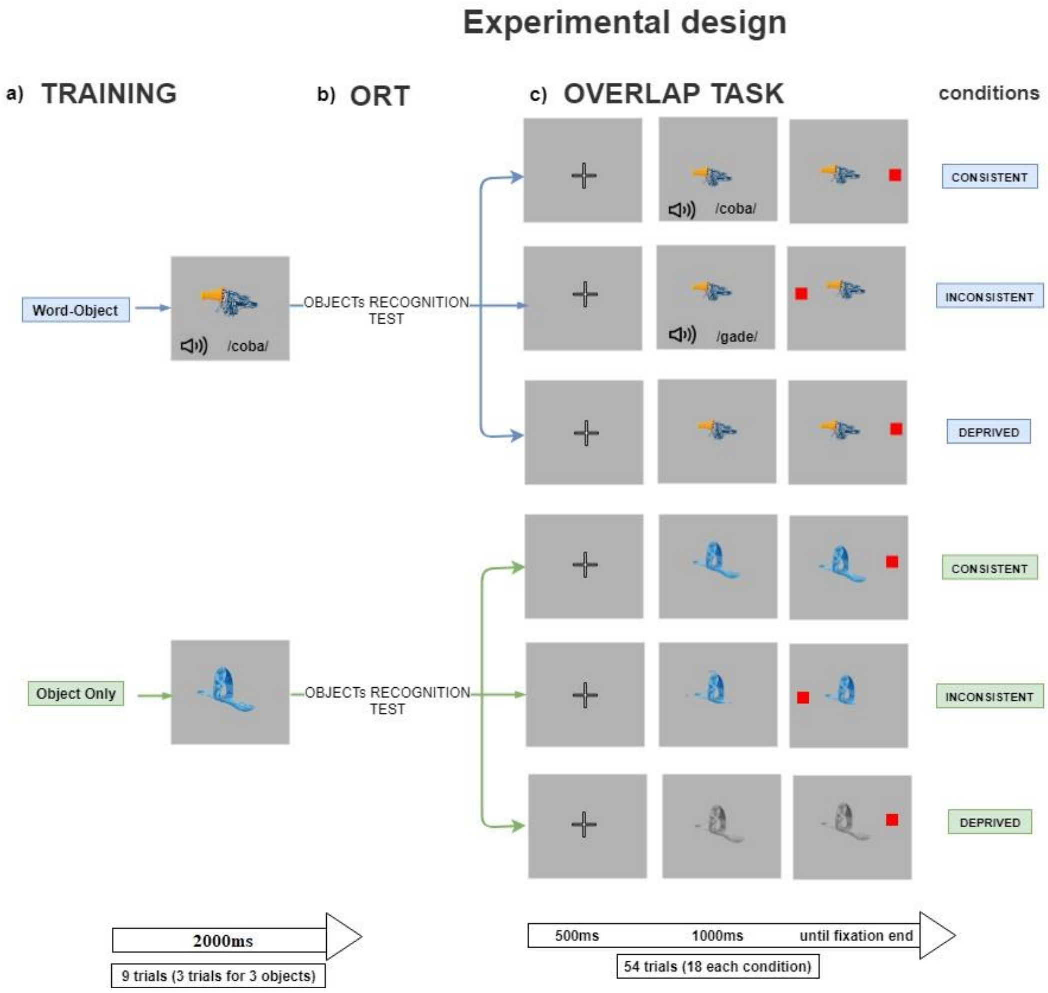


Figure 16. Participants were asked to pay attention to two a) training phases consisting of 9 trials (2 seconds each) in two different blocks (Word-Object and Object Only). Following each training participants performed the b) ORt, if they reached at least 80% of accuracy then they performed c) the Overlap task in which after a fixation cross (500 ms) a central object (S1) was presented in three conditions (Consistent, Inconsistent and Deprived) for 1 second, then a peripheral stimulus (S2) could randomly appear on the right or on the left of S1, at the same degree of eccentricity.

Calibration phase: Participants were given instruction about the calibration phase in which five markers were presented one by one in different locations on the screen (top-left, top-right, centre, bottom-left and bottom-right).

Training phase. Each participant was exposed to two trainings, the Object Only and the Word-Object. In the Object Only training, three static visual objects were presented one by one at the centre of the screen. Participants had to fixate each object (three randomised trials each) for ~2 seconds, to complete the training. In the Word-Object training, three static visual objects were presented one by one at the centre of the screen paired with, as soon as the participant fixated it, a pseudowords each (i.e., three randomised trials for three WO, ~2 seconds each). Overall, six objects (three objects each training) were shown counterbalanced between trainings and participants across data collection.

Objects Recognition task: This phase immediately followed each training phase. Participants were tested in the Object Recognition tasks (ORt) after the Word-Object and the Object-Only training, respectively. Each ORt involved 18 trials, showing the three trained objects and three new objects, in a random order. The new objects and new words were selected from the NOUN database. After the presentation of a fixation point for 1000 ms, a stimulus appeared centred on the screen and participants had up to 5 seconds to categorise it either as seen or unseen during the previous training. Responses were given pressing a button (z or m, counterbalanced among participants).

Overlap task. After each Training participants performed an Overlap tasks. It consisted of three conditions presented randomly namely, the Consistent, the Inconsistent and the Deprived condition. In the Consistent condition, the object was presented identical to the training phase. In the Inconsistent condition, the object showed an altered feature i.e. a novel label or novel shape. In the Deprived condition, the object was deprived of its label or colour. In all conditions, after 500ms of looking at the fixation cross, participants were asked to fixate the central object (S1) and after 1000 milliseconds of looking a peripheral stimulus (S2 i.e., a red square) appeared. Participants were instructed to end their fixation on S1 and reach S2 as soon as possible. S2 was randomly presented either on the left or the right of S1 (5.9° distance in visual degrees from the centre to the peripheral stimuli). The trial ended when participants moved their gaze away from the central stimulus, after the appearance of the peripheral stimulus. The three overlap conditions, i.e. Consistent, Inconsistent and Deprived, were presented in a random order. Each Overlap task comprised 54 trials (18 trials for each condition).

Apparatus

The visual stimuli were presented with Open Sesame software 3.1 (Psychology Software Tools, Pittsburgh, PA) on a 27-inch monitor (Philips 300x300). A remote, infrared eye-tracking camera (Tobii X2-60 Eye-Tracker) placed directly below the screen recorded the 60cm away participant's eye movements using bright-pupil technology at a sampling frequency of 60 Hz. The audio stimuli were recorded with Audacity software (Audacity Team, 2017; equipment: microphone SHURE PG58 and sound card M-AUDIO Fast Track), by a female Italian native speaker. The audio stimuli were presented with two speakers (KRK rokit rp 5) placed on the right and left of the screen. The experimental session took place in a room with semi-darkness constant luminance guaranteed by a lamp positioned one meter away behind the participant. The room presented a dark curtain that separated the participant area from the experimenter area.

Statistical Analysis

Data from the training phase and the overlap task were analysed including participants who reached at least 80% accuracy during the OR task. Descriptive analysis showed higher accuracy for the OO (M = 98, SD = 2.8) compared to the Word-Object (M = 95, SD = 5.8).

Saccadic latency the time interval between the appearance of S2 and the beginning of the saccadic movement from S1 towards it, during the overlap task (i.e., fixation end). Saccadic latency was defined as beginning within 0.1 and 1.0 seconds from the onset of S2 (Kenward et al., 2017).

Pupil dilation Pupil dilation variability consists of a tonic state and a phasic response (Hepach & Westermann, 2016), and only the latter was analysed. To obtain a measure of the phasic response, it was followed the Hepach and Westermann (2016) procedure by computing the median of the first 100 ms of epoch at the beginning of each trial – this constituted the baseline period – and subtracting them to the remaining of the trial. During the training, pupil dilation reflected stimuli encoding (Kucewicz et al., 2018; Attar, Schneps & Pomplun, 2013). During the Overlap task, it was analysed the time window between 100 and 1000 ms after S2 onset and before saccades to occur, to get a measure of deallocation of attention resources (Privitera et al., 2008). Note that in both cases, slower, but more prolonged dilation before eye movements, reflects more sustained increases in arousal or mental effort (Mathôt, 2018).

Both disengagement latencies and pupil dilation were analysed by means of generalised linear models (GzLMs) that allow me to account for both random and fixed effects. Note that GzLMs are an extension of the GLMs, that allow to specify the distribution family (e.g., gamma). When needed, this allows to overcome the assumptions made by GLMs that require residuals to be normally distributed, and variability is constant across the levels of the predictors (Fox, 2008). Since when dealing with non-negative behavioural data (e.g., disengagement latencies), residuals are often distributed with positive skewness and heteroscedasticity. Thus, the distribution family was chosen based on residual distributions. All models were fitted with the lme4 package (Bates, Maechler, Bolker, & Walker, 2015) in R software (Version 5.3.8., R Core Team, 2008). To find the best approximation to the true model, it was followed a model comparison approach, using likelihood ratio test (LTR), AIC (Akaike Information Criterion; Akaike, 1974) and AIC weight as indexes of goodness of fit. The former tests the hypothesis of no difference between the likelihoods of two nested models. The AIC and AIC weight give information on the models' relative evidence (i.e., likelihood and parsimony) so that the model with the lowest AIC and the highest AIC weight value is preferred (Wagenmakers & Farrell, 2004). The model comparison started from the simplest model that explains a smaller portion of variance and proceeded by adding predictors (Pitt, Myung, & Zhang, 2002). For each dependent variable – i.e., disengagement latencies and pupil dilation – five different models were tested. The model comparison started with the null model including only the random intercepts of participant (dependent variable ~ (1|Participant) + residuals); in the second model, it was introduced Training (Object Only vs. WordObject) as a fixed factor (dependent variable ~ training + (1|Participants) + residuals); in the third model, it was introduced the second fixed factor, i.e., overlap task (Consistent, Inconsistent and Deprived stimulus, (dependent variable ~ training + overlap task condition + (1|Participants) + residuals); the fourth model also included the interaction between the two fixed factors (dependent variable ~ training * overlap task condition + (1|Participants) + residuals). Outliers were evaluated by the Influence analysis for Generalized Mixed-Effects models (Nieuwenhuis, Grotenhuis & Pelzer, 2012).

4.2.2. Results

Pupillometry

Only data from both eyes and those observations that had a cook's distance smaller than 4 times the mean were analyzed, for further details see Appendix. Figure 17 shows that pupil dilation and then resource allocation, differently increased during the two Trainings,

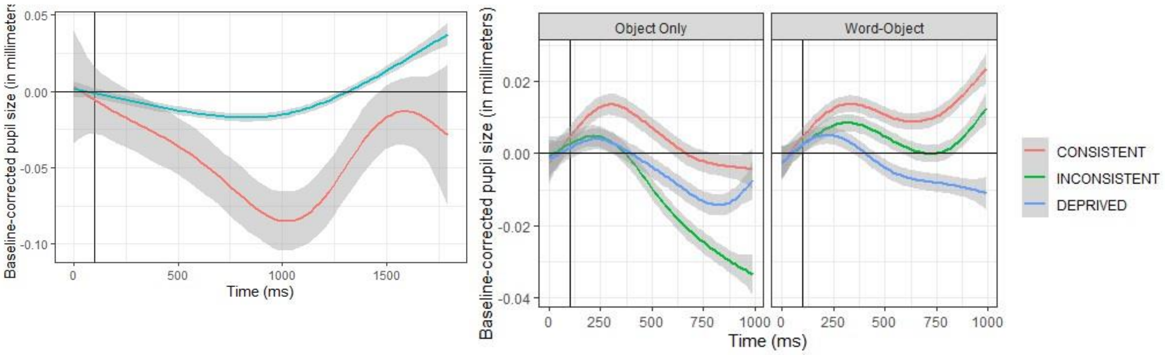
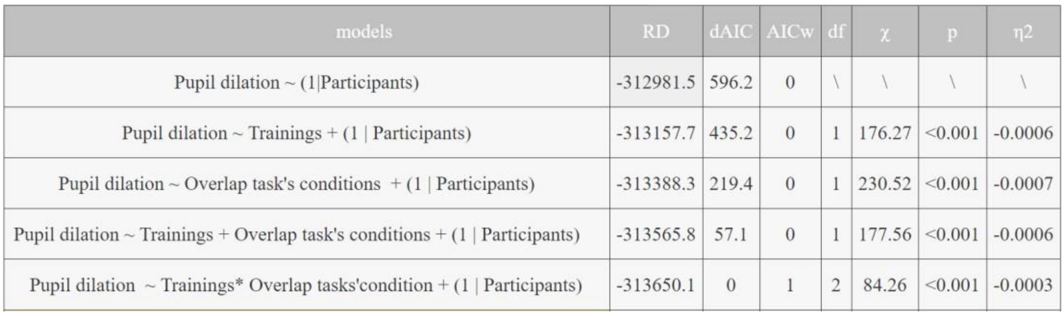


Figure 17. Baseline-corrected pupil size in millimetres across Time in milliseconds during both the Object Only (red) and the Word-Object (green) Training in the left panel; and during the Overlap task' conditions (Consistent, Inconsistent and Deprived) after the Object Only and the Word-Object training, on the right panel. Plotted the max and minimum values, the mean effect (bold line) and the 95% Confidence Intervals (CI, dark grey).

As shown in Table 3 the model comparison of the GzLMs with Gamma family, log link function and subject as random intercept showed the interactive model Training X Overlap tasks' conditions to better explain pupil dilation on S1 after S2 onset (before saccades to occur), during the Overlap task.



models	RD	dAIC	AICw	df	χ	p	η2
Pupil dilation ~ (1|Participants)	-312981.5	596.2	0	\	\	\	\
Pupil dilation ~ Trainings + (1 | Participants)	-313157.7	435.2	0	1	176.27	<0.001	-0.0006
Pupil dilation ~ Overlap task's conditions + (1 | Participants)	-313388.3	219.4	0	1	230.52	<0.001	-0.0007
Pupil dilation ~ Trainings + Overlap task's conditions + (1 | Participants)	-313565.8	57.1	0	1	177.56	<0.001	-0.0006
Pupil dilation ~ Trainings* Overlap tasks'condition + (1 | Participants)	-313650.1	0	1	2	84.26	<0.001	-0.0003

Table 3. Model comparison for predicting Pupil dilation in the Training phase. RD = residual deviance, AIC = Aikake information criterion, dAIC = difference between a model's AIC and those of the best model, AICw = AIC weight, df = degrees of freedom of the chi-squared statistic, η2 = eta squared as the ratio between the chi-squared and the residual deviance of the null model.

The main effect of Trainings showed higher pupil dilation in the Word-Object compared to the Object Only both during the training (β = 0.67, SE = 0.04, t = 14.95) and during the overlap task (β = 0.007, SE = 0.0005, t = 13.28). The interaction revealed significantly higher peak of participants' focused attention on S1 in the Consistent compared to both the Inconsistent (β = 0.009, SE = 0.0006, t= -15.47) and the Deprived (β = -0.01, SE = 0.0006, t = -18.99) conditions. These results suggest the presence of a consistency effect that is, presenting S1 without a feature or with a novel feature required less resources allocation on S1 during disengagement of attention. Furthermore, the interaction revealed two opposite pattern in the Object Only and the Word-Object block, respectively. After the Object Only block, the Inconsistent condition i.e. novel shape, showed a significant pupil restriction compared to the Deprived condition i.e. no colour (β =-0.01, SE = 0.001, t =

11.1). Whereas, after the Word-Object block the Inconsistent condition i.e. novel label, showed significantly higher pupil dilation compared to the Deprived condition i.e., no label (β =-0.01, SE = 0.001, t = -11.1).

Saccadic latency

During the overlap task participants were presented with three conditions namely, the Consistent condition when S1 was presented identical to the previous training, the Inconsistent condition when the object was paired with a novel word or a novel shape; and the Deprived condition when S1 was presented with no label or no colour in the Word-Object (WO) pair and the Object Only (OO), respectively. The time window analysed corresponded to the epoch before participants ended their fixation at S1 after S2 onset (>0.1 and <1 second). Figure 18 shows the mean and the standard errors for each condition of the Object Only and Word-Object blocks.

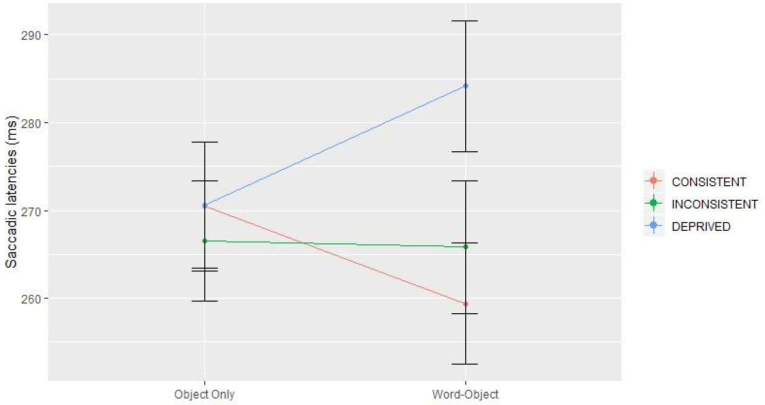
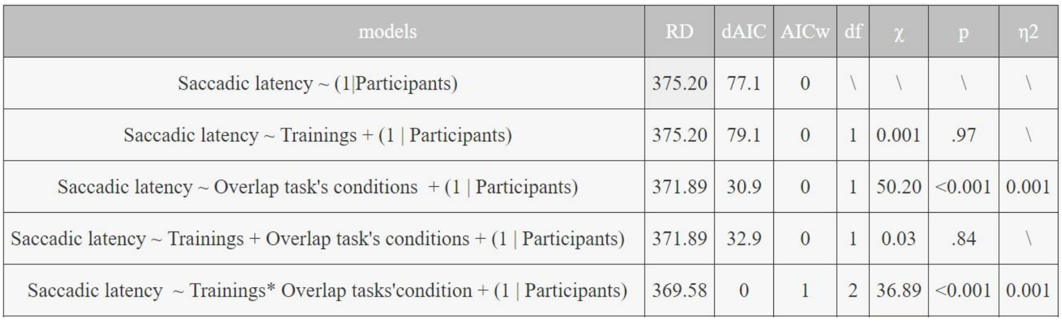


Figure 18. Interaction plot of Saccadic latency during the Overlap task's conditions further split for Object Only and Word-Object. Mean values and SE are plotted.

As shown in Table 4, models comparison of GzLMs with Gamma family, identity link function and subjects as random intercepts showed the interactive model to explain the saccadic latency measure better. The Training emerged not to be a good predictor of saccadic latency during the Overlap task (β = 0.0001, SE = 0.006, t = -0.02). The interaction showed that in the Word-Object block, the Consistent condition showed significantly shorter saccadic latency compared to the Deprived i.e. no label, condition (β = 0.05, SE = 0.008, t= 6.23). The Deprived condition showed longer latency compared to Inconsistent condition (β = 0.06, SE = 0.02, t = 2.29). No differences emerged comparing the Consistent to the Inconsistent condition i.e. novel label (β = 0.005, SE = 0.008, t =0.7). No difference among conditions emerged by comparing the Conditions in the Object Only block.



models	RD	dAIC	AICw	df	χ	p	η2
Saccadic latency ~ (1|Participants)	375.20	77.1	0	\	\	\	\
Saccadic latency ~ Trainings + (1 | Participants)	375.20	79.1	0	1	0.001	.97	\
Saccadic latency ~ Overlap task's conditions + (1 | Participants)	371.89	30.9	0	1	50.20	<0.001	0.001
Saccadic latency ~ Trainings + Overlap task's conditions + (1 | Participants)	371.89	32.9	0	1	0.03	.84	\
Saccadic latency ~ Trainings* Overlap tasks'condition + (1 | Participants)	369.58	0	1	2	36.89	<0.001	0.001

Table 4. Model comparison for predicting Saccadic latency during the Overlap tasks' condition after both the Object Only and the Word Object training. RD = residual deviance, AIC = Aikake information criterion, dAIC = difference between a model's AIC and those of the best model, AICw = AIC weight, df = degrees of freedom of the chi-squared statistic, η2 = eta squared as the ratio between the chi-squared and the residual deviance of the null model.

On-line and after- effect of words

In order to test any on-line word facilitation effect, it was calculated the difference between saccadic latency of the Consistent conditions of the Object Only block and the Consistent condition of the Word-Object block. The afterword effect i.e. the effect due to the representations triggered by each training, was calculated as the difference between the saccadic latency in the Consistent condition of the Object Only block and the Deprived condition of the Word-Object (physically identical) block. Positive values suggest an on-line and/or afterword facilitation effect, see Figure 19.

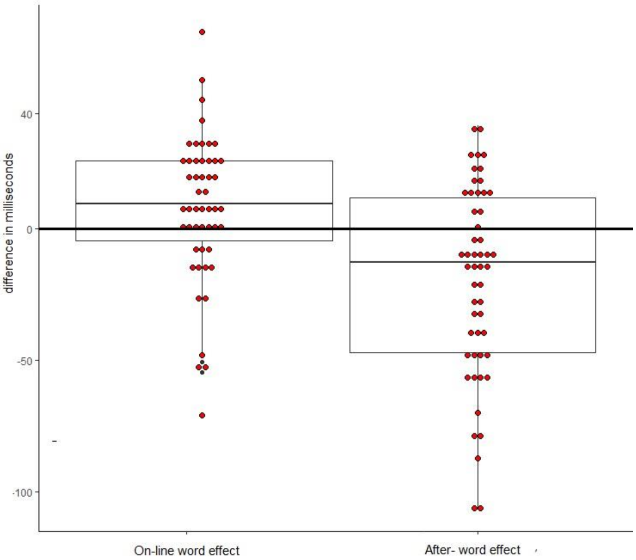


Figure 19. Box plot and dot plot of the computed On-line word effect (Consistent Object Only – Consistent Word-Object) and Afterword effect (Consistent Object Only – Deprived Word-Object) in milliseconds, for each subject. First, second and third quartile are plotted. The black line indicates the facilitation effect of word (>0).

Table 5 shows the model comparison suggesting that the online word effect was significantly larger compared to the after- word effect (β = -25.49, SE = 5.51, t = -4.63)

models	RD	dAIC	AICw	df	χ	p	η2
ð saccadic latency ~ (1|Participants)	99	20.9	0	\	\	\	\
ð saccadic latency ~ On-line vs After word effect + (1 | Participants)	98	0	1	1	17.65	<0.0001	0.01

Table 5. Model comparison for predicting difference (∂) in Saccadic latency by the on-line word effect (Consistent OO – Consistent WO) and the after-word effect (Consistent OO – Deprived WO). RD = residual deviance, AIC = Aikake information criterion, dAIC = difference between a model's AIC and those of the best model, AICw = AIC weight, df = degrees of freedom of the chi-squared statistic, η2 = eta squared as the ratio between the chi-squared and the residual deviance of the null model.

4.2.3. Discussion experiment 1

This study investigated whether and to what extent stored representations influence disengagement of visual attention from familiar objects. First, the results from the training showed that individuals deployed higher attention resources to Word-Object pairs than Object Only during information encoding in the two trainings. However, it might be the case that the two different kinds of information (audio-video vs video only) rather than linguistic information *per se*, led to different pupil phasic response in the two trainings. In fact, multi-sensory integration is thought to recruit higher attentional resources compared to unimodal information processing by triggering the learning processes (Gogate, Walker-Andrews, & Bahrick, 2001). To disentangle these interpretations, in Experiment 2 participants were asked to pay attention to two auditory conditions: Word-Object (WO) and Tone-Object (TO) pairs, in two separate blocks.

Second, pupil dilation is usually associated with inhibitory control during disengagement of attention by showing larger preparatory dilations predicting faster subsequent saccades, an index of covert shift of attention preceding each saccadic eye movement (Deubel & Schneider, 1996; Hoffman & Subramaniam, 1995; Kowler, Anderson, Dosher, & Blaser, 1995). Here, the consistent on-line label (Consistent condition in WO) prompted higher pupil dilation and facilitated disengagement of attention compared to the consistent OO (Consistent condition in OO). Importantly, the comparison between the on-line and the afterword presentation showed a larger facilitatory effect due to the on-line label presentation, in line to what expected by the Label feedback hypothesis. Furthermore, Experiment 1 corroborated and extended Lupyan and Ward's (2013) findings, by showing that the presence of a spoken label boosted online object selection and facilitated disengagement of attention mechanism.

However, in contrast with Lupyan and Ward (2013), the visual object with no label (Deprived condition in WO) did not efficiently trigger disengagement of attention by lowering down saccadic latency during the Overlap task, compared to presence of an inconsistent label (Inconsistent condition after WO). This finding might be explained by the fact that, contrarily to the active object recognition tested by the authors (participants should explicitly identify the object), this study investigated stimulus-driven mechanisms in a simple spatial task. Hence, it is possible that during such a task the auditory presentation of a label did trigger deployment of attention from the object more efficiently compared to the silent presentation of visual stimuli (off-line label representation). Thus, any top-down effect due to the Word-Object training might be confounded due to this specific design that did reduce reference uncertainty (a single word-object pair) making easier for the participants to perform the task with any label compared to the no-label condition, like in Experiment 1. However, to disambiguate any bottom-up facilitation (multisensory vs visual) and concurrently, to investigate any top-down linguistic effect during such a task in Experiment 2 participants performed a similar task paying attention to two multisensory conditions: Word-Object and Tone-Object pairs i.e. sinewave transformed tones from pseudowords, in which S1 could be presented in a Consistent and in an Inconsistent condition.

4.2.4. Method

The experiment 2 was identical to experiment 1 apart from the Object Only training that was substituted by the Tone-Object (TO) training in which 6 novel objects (see Stimuli in the Method of Experiment 1) paired with 6 tone each, were presented. The tones were sinewave transformation matched in amplitude and duration to the pseudowords (Fulkerson and Waxman, 2007, see Appendix).

In the overlap task, two conditions were presented, namely, a Consistent i.e. S1 identical, and Inconsistent condition that showed S1 with a different word or tone paired with the familiar visual stimulus.

Participants

Fifty-two adults (35 female, mean age = 24, SD = 2.4,) were enrolled. The inclusion criteria for all participants were to be in good health, to have no sensorial or neurological disorders or any familial language disorder. Two participants were excluded from the final analysis because of technical issues during the experimental session. The ORt showed higher accuracy in Tone-Object (M = 83, SD = 6.3) compared the Word-Object block (M=93, SD = 6.9).

Pupillometry

Figure 20 show baseline-corrected pupil dilation (in millimetres) across the time of the averaged trial (50 participants, 9 trials, 2 seconds each) during Word-Object and Tone-Object trainings and during the Overlap task' conditions: Consistent in red and Inconsistent in green, after the Tone and the Word-Object training.

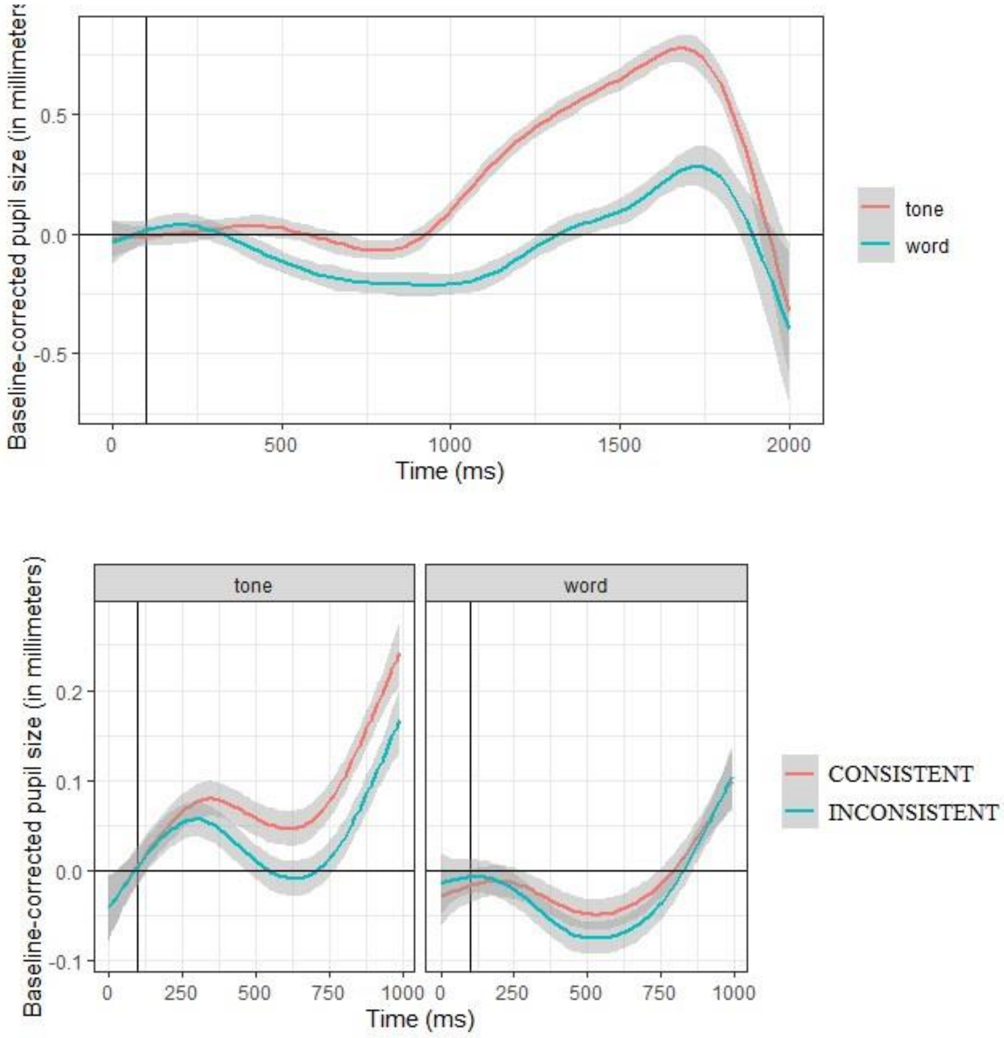
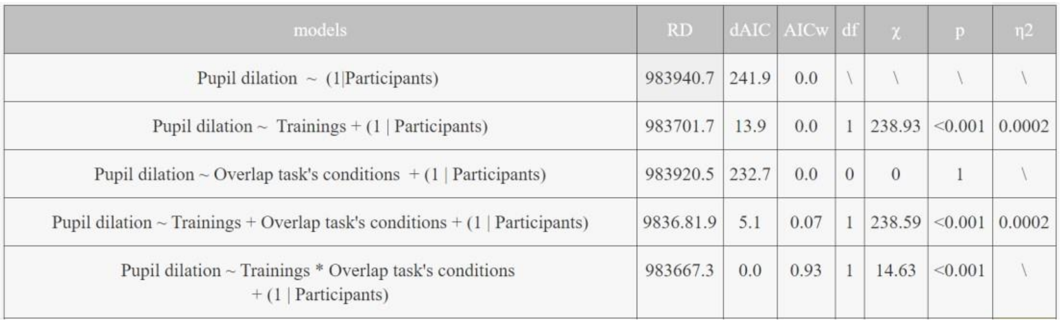


Figure 20. Baseline-corrected pupil size in millimetres across Time in milliseconds during each trial of both the Tone (red) and the Word (green) training and across the Overlap task's conditions (Consistent and Inconsistent) in the Tone-Object and the Word-Object block. Plotted the mean effect (bold line) and the 95% Confidence Intervals (CI, dark grey).

Table 6 shows the model comparison predicting baseline-corrected pupil dilation in the time window after S2 onset and before saccadic latency to occur, during the Consistent and Inconsistent condition in the Overlap task for the Tone-Object and the Word-Object blocks, respectively. The interactive model Training x Conditions emerged to better approximate the data.



models	RD	dAIC	AICw	df	χ	p	η2
Pupil dilation ~ (1|Participants)	983940.7	241.9	0.0	\	\	\	\
Pupil dilation ~ Trainings + (1 | Participants)	983701.7	13.9	0.0	1	238.93	<0.001	0.0002
Pupil dilation ~ Overlap task's conditions + (1 | Participants)	983920.5	232.7	0.0	0	0	1	\
Pupil dilation ~ Trainings + Overlap task's conditions + (1 | Participants)	9836.81.9	5.1	0.07	1	238.59	<0.001	0.0002
Pupil dilation ~ Trainings * Overlap task's conditions + (1 | Participants)	983667.3	0.0	0.93	1	14.63	<0.001	\

Table 6. Model comparison for predicting Pupil dilation in millimetre during the Overlap task's conditions in the Tone-Object and the Word-Object blocks. RD = residual deviance, AIC = Aikake information criterion, dAIC = difference between a model's AIC and those of the best model, AICw = AIC weight, df = degrees of freedom of the chi-squared statistic, η2 = eta squared as the ratio between the chi-squared and the residual deviance of the null model.

The interactive models showed the Tone-Object training to lead to higher and sustained pupil dilation compared to the Word-Object training (β = 0.09, SE = 0.006, t = 13.70). The interaction showed a consistency facilitatory effect but only in the Tone-Object block (β = 0.04, SE =0.006, t = 5.86) whereas it did not predicted any effect due to the label consistency in the Word-Object block (β = -0.006, SE = 0.006, t= -0.93).

Saccadic latency

Figure 21 shows the saccadic latency's mean and the standard error (SE) in the two conditions (Consistent and Inconsistent) of the Overlap in the Word-Object and the Tone-Object blocks.

The model comparison in Table 7 shows that the model that better approximated our data was the additive model with the single fixed factor Training. More specifically, shorter saccadic latency emerged in the Word-Object compared to the Tone-Object block (β = 6.33, SE = 3.44, t = 1.84). Whereas no main effect among the Overlap tasks' condition emerged (β = 0.81, SE = 3.44, t = 0.23).

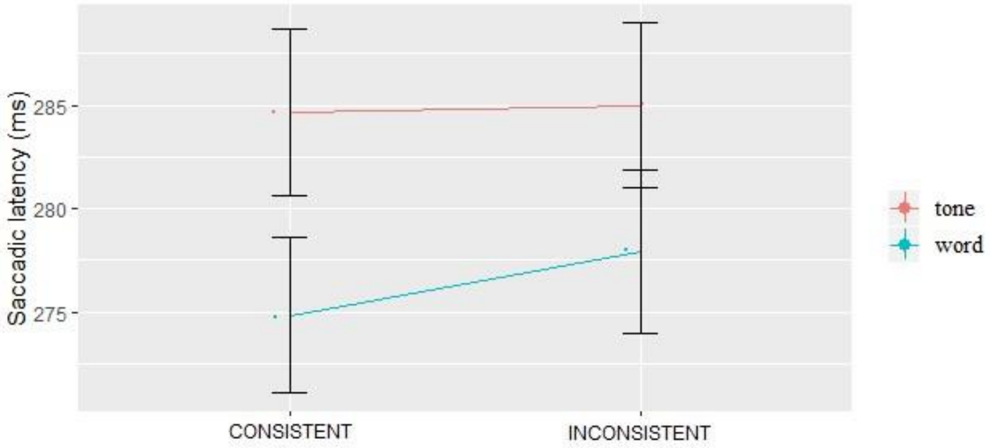


Figure 21. Interaction plot of Saccadic latency during the Overlap task's conditions further split for Tone-Object and Word-Object blocks. Mean values and SE are plotted.

models	RD	dAIC	AICw	df	χ	p	η2
Saccadic latency ~ (1|Participants)	5112	2.0	0.17	\	\	\	\
Saccadic latency ~ Trainings + (1 | Participants)	5111	0.0	0.50	1	4.04	<0.04	0.001
Saccadic latency ~ Overlap task's conditions + (1 | Participants)	5111	4.0	0.07	0	0	1	\
Saccadic latency ~ Trainings + Overlap task's conditions + (1 | Participants)	5110	1.9	0.19	1	4.03	<0.04	0.001
Saccadic latency ~ Trainings * Overlap task's conditions + (1 | Participants)	5109	3.9	0.07	1	0	1	\

Table 7. Model comparison for predicting Saccadic latency in milliseconds during the Overlap task's conditions in the Tone-Object and in the Word-Object blocks. RD = residual deviance, AIC = Aikake information criterion, dAIC = difference between a model's AIC and those of the best model, AICw = AIC weight, df = degrees of freedom of the chi-squared statistic, η2 = eta squared as the ratio between the chi-squared and the residual deviance of the null model.

4.2.5. Discussion experiment 2

Experiment two was set to disambiguate the facilitatory label effect on attention mechanisms from a general auditory and/or intersensory facilitation during such a spatial task. The analysis of saccadic latency in experiment 2 did show a facilitatory effect on disengagement of attention from S1 driven by linguistic labels rather than on auditory information per se (Lupyan & Thompson-schill, 2012). In addition, pupil phasic variation during the two separate trainings did show that the Tone-Object pairs differently affected stimulus encoding compared to Word-Objects. Indeed, Tone-Objects did require more attentional effort and resource allocation to be encoded and it impaired representation encoding as shown by the lower accuracy score in the ORt, participants performed that right after the Tone-Object and the Word-Object training. Finally, the two trainings successfully disambiguated the presence of different mechanisms operating during the Overlap task's conditions, depending on the training manipulation. The resources recruited during the Overlap task in the Tone-Object stayed higher compared those predicted by the Word-Object block, by suggesting a bottom-up interference lead by multisensory stimulation. On the contrary, the Word-Object pair seems to facilitate via top-down mechanisms the encoding and the representation in memory of the stimulus and did predict faster saccadic latency in the subsequent spatial task. Finally, a consistency effect did emerge in the Tone-Object in which a valid pair triggered higher resource allocation before visual disengagement occurred, compared to the invalid Tone-Object pair. Whereas any consistency effect did not emerge for Word-Object pairs.

4.2.6. GENERAL DISCUSSION

Language can affect perceptual representations (Lupyan, 2012) and facilitate orienting of attention. It has already been shown, that words prompt object recognition (Lupyan & Ward, 2013) by helping participants to selectively represent and in turn, efficiently identify a target (Wolfe & Horowitz, 2004; Eimer, 1996). However, less is known about the effect of linguistic and top-down information on attention deployment guided by stimulus-driven mechanisms during simple spatial tasks (Zivony & Lamy, 2016). This study investigated the effect of top-down information about Word-Objects, Tone-Objects and Object-Only on early mechanisms of attention during a traditional spatial task.

In Experiment 1 even when linguistic cue constituted entirely redundant information because participants already knew what they should do on each randomly intermixed trial, i.e. fixate S1 and then reach S2 as soon as possible, they disengaged their attention more quickly during the consistent label presentation, in both experiments. This type of facilitation occurred before attention shift, as shown by higher pupil phasic response predicting shorter latency of eye movements, in Experiment 1. The facilitation due to hearing a label was carried through the entire experiment 1 yet, the difference between the intermixed label and no-label trials did persisted (Lupyan and Spivey, 2010b) which was only possible if hearing a label affected perceptual processing in a transient, on-line manner (Experiment 1). A reversed effect did emerge with more complex items such as Tone-Object pairs that elicited a more intense phasic response; hence, recruited higher resource allocation to de-allocate attention but concurrently, slowed down disengagement of attention. These findings successfully disambiguate the facilitatory effect found across the two experiments that should depend on top-down mechanisms selectively operating over linguistic labels.

Furthermore, the consistency effect found across the two experiments suggests that the mental representation of the object can be activated in a memory-based fashion, during simple spatial tasks (Zivony & Lamy, 2016). This specific task should mainly recruit stimulus-driven mechanisms. Nevertheless, the facilitatory top-down effect of label on disengagement of attention from familiar stimuli might be prompted by the reduced reference uncertainty across the two experiments (a single word-object pair). This makes this paradigm particularly suitable to investigate linguistic influence on perceptual representations and on pure spatial mechanisms. Finally, the facilitatory effect prompted by the label, independently from any condition in experiment 2, suggest the possibility that linguistic representations and not simply multisensory representation do allow the retrieval of more stable reference frameworks that efficiently guide object identification by reducing its perceptual ambiguity even in the presence of inconsistent labels and subsequently do trigger perceptual representation and visual orienting of attention.

This study aimed to bridge different approach to understand the same act of cognition that is identifying visual objects and orienting attention in space depending on the auditory association that can co-occur in the environment. The concomitant eye-movements and phasic pupil measures allowed to disambiguate from automatic nervous responses and oculomotor mediated response to visual stimuli depending on previous exposure; thus this study offer a promising insight into mental representation prompted by linguistic and auditory cues that may are differently affected by both bottom-up interference and top-down guidance.

Limitation and further perspectives

A major problem of this study relies on statistical power, in fact even if by following the rule-of-thumb (N = 10 for each parameter of the multiplicative model) (Austin & Steyerberg, 2015) the two experiments (N = 115) still resulted somewhat underpowered. Thus, further replications should confirm the directions and the size of such effects found here. Hopefully, the effort to use free sources: database (stimuli), software for experiments (Opensesame) and software for the analysis (R free software) should facilitate a precise replication of the same o similar investigation, across laboratories. To conclude, this study offered insightful information about mental representation encoding and linguistic-attentive dynamics in adults; further studies should clarify if the same pattern of outcomes do emerge in a context with higher referential complexity (more visual-auditory pairs) and linguistic complexity i.e. sentences.

4.2.7. Bibliography

Attar, N., Schneps, M., & Pomplun, M. (2013). Pupil size as a measure of working memory load during a complex visual search task. *Journal of Vision*, 13*(9)*, 160160.

Austin, P. C., & Steyerberg, E. W. (2015). The number of subjects per variable required in linear regression analyses. *Journal of clinical epidemiology*, 68*(6)*, 627636.

Bahrick, L. E., & Lickliter, R. (2000). Intersensory Redundancy Guides Attentional Selectivity and Perceptual Learning in Infancy. 36(2), 190–201.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1–7. 2014.

Deubel, H., & Schneider, W. X. (1996). Saccade target selection and object recognition: Evidence for a common attentional mechanism. *Vision research*, 36*(12)*, 1827-1837.

Edmiston, P., & Lupyan, G. (2015). What makes words special ? Words as unmotivated cues. Cognition, 143, 93–100.
https://doi.org/10.1016/j.cognition.2015.06.008

Eimer, M. (1996). The N2pc component as an indicator of attentional selectivity. *Electroencephalography and clinical neurophysiology*, 99*(3)*, 225-234.

Fischer, B., & Breitmeyer, B. (1987). Mechanisms of visual attention revealed by saccadic eye movements. *Neuropsychologia*, 25*(1)*, 73-83.

Fulkerson, A. L., & Waxman, S. R. (2007). Words (but not tones) facilitate object categorization: Evidence from 6-and 12-month-olds. *Cognition*, 105*(1)*, 218-228.

Gogate, L. J., Walker-Andrews, A. S., & Bahrick, L. E. (2001). The intersensory origins of word comprehension: an ecological–dynamic systems view. *Developmental Scienc*e, 4*(1)*, 1-18.

Hayhoe, M. M., Shrivastava, A., Mruczek, R., & Pelz, J. B. (2003). Visual memory and motor planning in a natural task. *Journal of vision*, 3*(1)*, 6-6.

Henderson, J. M., Malcolm, G. L., & Schandl, C. (2009). Searching in the dark: Cognitive relevance drives attention in real-world scenes. *Psychonomic bulletin & review*, 16*(5)*, 850-856.

Hepach, R., & Westermann, G. (2016). Pupillometry in infancy research. *Journal of Cognition and Development,* 17*(3)*, 359-377.

Hoffman, J. E., & Subramaniam, B. (1995). The role of visual attention in saccadic eye movements. *Perception & psychophysics*, 57*(6)*, 787-795.

Horst, J. S., & Hout, M. C. (2016). The Novel Object and Unusual Name (NOUN) Database: A collection of novel images for use in experimental research. *Behavior research methods*, 48(4), 1393-1409.

James, W. (1890). The principles of psychology. Harvard University Press.

Jovancevic, J., Sullivan, B., & Hayhoe, M. (2006). Control of attention and gaze in complex environments. *Journal of Vision*, 6*(12)*, 9-9.

Kowler, E., Anderson, E., Dosher, B., & Blaser, E. (1995). The role of attention in the programming of saccades. *Vision research*, 35*(13)*, 1897-1916.

Kucewicz, M. T., Dolezal, J., Kremen, V., Berry, B. M., Miller, L. R., Magee, A. L., ... & Worrell, G. A. (2018). Pupil size reflects successful encoding and recall of memory in humans. *Scientific reports*, 8(1), 4949.

Logan, G. D. (2002). An instance theory of attention and memory. Psychological Review, 109(2), 376.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, 390*(6657)*, 279.

Lupyan, G. (2012). Linguistically modulated perception and cognition: The labelfeedback hypothesis. Frontiers in Psychology, 3(MAR), 1–13.

Lupyan, G., & Spivey, M. J. (2007). Perceptual processing is facilitated by ascribing meaning to novel stimuli. Current Biology, 18(10), R410–R412. https://doi.org/10.1016/j.cub.2008.02.073

Lupyan, G., & Thompson-schill, S. L. (2012). The Evocative Power of Words : Activation of Concepts by Verbal and Nonverbal Means. 141(1), 170–186. https://doi.org/10.1037/a0024904

Lupyan, G., & Ward, E. J. (2013). Language can boost otherwise unseen objects into visual awareness. https://doi.org/10.1073/pnas.1303312110

Mathôt, S. (2018). Pupillometry: Psychology, physiology, and function. *Journal of Cognition*, 1*(1)*, 1-23. https://doi.org/10.5334/joc.18

Mazza, V., Turatto, M., Umiltà, C., & Eimer, M. (2007). Attentional selection and identification of visual objects are reflected by distinct electrophysiological responses*. Experimental Brain Research*, 181*(3)*, 531-536.

Morris, N., & Jones, D. M. (1990). Memory updating in working memory: The role of the central executive. *British journal of psychology*, 81*(2)*, 111-121.

Nako, R., Wu, R., Smith J., T. J., & Eimer, M. (2014). Item and category-based attentional control during search for real-world objects: Can you find the pants among the pans? Journal of Experimental Psychology: Human Perception and Performance, 40(4), 1283–1288. https://doi.org/10.1037/a0036885

Neider, M. B., & Zelinsky, G. J. (2006). Scene context guides eye movements during visual search. *Vision research*, 46*(5)*, 614-621.

Nieuwenhuis, R., te Grotenhuis, M., & Pelzer, B. (2012). Influence. ME: Tools for detecting influential data in mixed effects models. *R journal,* 4 *(2)*. 38–47.

Pitt, M. A., Myung, I. J., & Zhang, S. (2002). Toward a method of selecting among computational models of cognition. *Psychological review*, *109*(3), 472.

Privitera, C. M., Renninger, L. W., Carney, T., Klein, S., & Aguilar, M. (2008, February). The pupil dilation response to visual detection. In *Human Vision and Electronic*

Imaging XIII (Vol. 6806, p. 68060T). International Society for Optics and Photonics.

Rakison, D. H., Lupyan, G., Oakes, L. M., & Walker‐Andrews, A. S. (2008). Developing object concepts in infancy: An associative learning perspective. *Monographs of the Society for Research in Child Development*, i-127.

R Development Core Team (2008). R: A language and environment for statistical computing. Vienna, Austria. Available from: http://www.R-project.org

Saslow, M. G. (1967). Effects of components of displacement-step stimuli upon latency for saccadic eye movement. *Josa*, 57*(8)*, 1024-1029.

Spinelli, G., Sulpizio, S., & Burani, C. (2017). Q2Stress: A database for multiple cues to stress assignment in Italian. *Behavior research methods*, 49*(6)*, 2113-2126.

Tatler, B. W., Wade, N. J., Kwan, H., Findlay, J. M., & Velichkovsky, B. M. (2010). Yarbus, eye movements, and vision. *i-Perception*, 1*(1)*, 7-27.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive psychology*, 12*(1)*, 97-136.

Van der Stigchel, S., Hessels, R. S., van Elst, J. C., & Kemner, C. (2017). The disengagement of visual attention in the gap paradigm across adolescence. *Experimental Brain Research*, 235(12), 3585-3592.

Van Steenbergen, H., & Band, G. P. (2013). Pupil dilation in the Simon task as a marker of conflict processing. *Frontiers in human neuroscience*, 7, 215.

Wagenmakers, E. J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review*, 11(1), 192-196.

Wang, C. A., Boehnke, S. E., White, B. J., & Munoz, D. P. (2012). Microstimulation of the monkey superior colliculus induces pupil dilation without evoking saccades. *Journal of Neuroscience*, 32*(11)*, 3629-3636

Wolfe, J. M., & Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it? Nature Reviews Neuroscience, 5(6), 495.

Zivony, A., & Lamy, D. (2016). Attentional capture and engagement during the attentional blink: A "camera" metaphor of attention. Journal of Experimental Psychology: Human Perception and Performance, 42(11), 1886.

4.3. Target representation during category-based visual search: is there any word advantage?[6]

Do you have anything to take note? In our every-day environment, we can search for any of several items instead of searching for an exact, specific item. During such a visual search, the brain is actively separating similar perceptual stimuli into different categories, whereas it is grouping different perceptual stimuli into similar categories (Freedman, 2001). It is the interplay between the category classification cued by language (e.g., *anything to take a note* = stationery) and the diagnostic features of the visible objects (e.g., colour and shape) that guides our target selection among distractor items (e.g. pens among chopsticks) (Goldstone, 1994b; Goldstone, Kersten, & Carvalho, 2018; Harnad, 1987; Josephs & Konkle, 2017).

The functional hypothesis underpinning that active search mechanisms expects attention deployment to be template-guided during target selection. Classically, active search requires individuals to attend to a target before search initiation. Then, the visual features of the target are maintained in working-memory and organised in a detailed attentional template (Duncan & Humphreys, 1989; Olivers, Peters, Houtkamp, & Roelfsema, 2011; Wolfe & Horowitz, 2004). The template guided search starts as soon as the onset of the search array and it ends with the selection of the target, in a spatial fashion (Desimone & Duncan, 1995; Eimer, 2014; Eimer & Grubert, 2014). The time

---

[6] This study has been carried out in collaboration with Rebecca Nako and Martin Eimer.
Data were collected at the Brain and Behaviour Lab, Centre for Brain and Cognitive Development Department of Psychological Sciences, Birkbeck College University of London.

course of target selection has been widely investigated through the N2pc, a negative ERP component (180-300 milliseconds after the search display onset) that shows its maximal amplitude at posterior electrodes contralateral to the position of the target (Luck & Hillyard, 1994). The N2pc reflects target selection (Eimer, 1996; Mazza, Turatto, & Caramazza, 2009) and its amplitude varies as a function of the match between the mental representation (attentional template) and the perceptual features of the target (Nako, Smith, & Eimer, 2004; Nako, Wu, Smith & Eimer, 2014; Schneider, Beste, & Wascher, 2012). Such search guidance relies on the active recollection of the diagnostic features of an exact target (Wolfe & Horowitz, 2004), whereas in category-search tasks the visual diagnostic features of the target are not depicted in any cue (Maxfield, Stalder, Brook, & Zelinsky, 2014).

Moreover, defining the list of items that unambiguously fulfil a category is not an easy task when performed with basic-level category-items sharing more visual (perceptual distinctiveness) and semantic features (conceptual distinctiveness) (Konkle, Brady, & Alvarez, 2010; Maxfield et al., 2014). Furthermore, it becomes even more difficult by considering high-level category-items that are non-canonical in appearance by requiring the maintenance of few visual features in working-memory (Blair, Watson, & Meier, 2009; Maxfield & Zelinsky, 2012). Thus, the template-guided search for category items should be impossible due to the under-specification of visual diagnostic features of the target in WM (Yang & Zelinsky, 2009)

Nonetheless, Cunningham, & Wolfe (2012) showed that actively searching for categorical items is possible, even if it is markedly slower than searching for specific items (search times for category-items were on average about 32 ms/item slower than for specific items). It has been shown that the N2pc that has been mainly studied by using specific items search can also be elicited by category cues (Jenkins, Grubert, & Eimer, 2018) yet it shows a smaller amplitude for pictorial category cues compared to exact-item cues. Nako et al. (2014) investigated category-attentional control mechanisms by comparing exact items (e.g., a T-shirt during search for a t-shirt) and category cues (e.g., a t-shirt and pants to search for pants), in a visual search task where stimuli were line drawings of real-world visual objects. The authors found that target selection was efficiently guided by categorical cues as shown by both accuracy and the N2pc, and they concluded that the attentional template could hold category representations while operating for multiple items (e.g. to search either for a t-shirt and pants). However, the perceptual attributes that delimitate the borders of each category are still unknown (Freedman, 2001) and it is unclear whether the N2pc elicited by categories reflects either category membership or physical similarity.

Lupyan, Rakison, and McClelland, (2007) asked participants to categorise several novel objects ("aliens") in two groups (i.e., aliens to avoid vs aliens to approach) and received a feedback after each trial. In the 50% of the trials, the feedback was paired with a nonsense label (either spoken or written). The findings showed that learning a verbal association between the alien object and the label did facilitate perceptual categorisation. This evidence suggested that it was the linguistic information that made perceptual categorisation less ambiguous.

In other six experiments, Lupyan and Spivey (2010) investigated the effect of hearing the redundant (noninformative) name of familiar items (e.g., hearing the name two while seeing the number 2) during a series of probe detection tasks. The authors presented in the search array two groups of numeric digits in two possible fonts (e.g. **2**, **2**, **5**, **5**). The search array was preceded by either an exact-item (e.g., 2) or category cues (e.g. two). The authors explored the time course of the label effect on probe detection, by parametrically varying the delay between the appearance of the search array and the appearance of the attentional probe. They found that despite the word cues carried redundant information, the name of an item prompted faster detection of attention probes near to the named item. Specifically, the effects of language facilitation on attention deployment was stronger for the canonical font than for non-canonical fonts(Maxfield et al., 2014; Nako, Wu, Smith, et al., 2014; E.

H. Rosch, 1973) suggesting that the representation of the category (e.g., two) was likely canonical (e.g. probe detected faster near to 2 compared to 2) (Konkle & Oliva, 2012; Lupyan & Spivey, 2010; Maxfield et al., 2014; Schmidt & Zelinsky, 2009b). Altogether these findings showed that hearing a label associated with an object did guide visual detection, even when eye movements and attention were constrained by rapidly presenting the search array (Lupyan & Spivey, 2010).

Moving a step forward, the present study investigated how template-guided category search is triggered by word and pictorial cues, depending on category classification (i.e., basic and supraordinate categories) and its further interaction with perceptual distinctiveness of the category-items (Konkle et al., 2010). Here, two separate experiments investigated template-guided search cued by picture and words, for items of basic-level categories (e.g., pen to search for a pen) and supraordinate categories (e.g., stationery to search for a pen), respectively. Specifically, I tried to disentangle the degree at which the perceptual (canonical and non-canonical shape) and the conceptual (basic and supraordinate level) features of the template cued by words influenced the target selection (Clarke, 2017). In both experiments, I expected the search efficiency to increase as a function of the match between the semantic and the perceptual categorisation. In both experiments, I investigated target selection during category-guided search through the measurement of the amplitude of the N2pc ERPs component. Larger N2pc amplitude was expected for word cues compared to picture cues depending on a function of conceptual distinctiveness of categories.

4.3.1. Methods

Participants

Twenty-two native English-speaking participants with normal or corrected to normal vision (13 females; mean age = 28 years old, SD = 9 years) took part in the experiment one and were reimbursed for their participation. All participants signed informed consent, and the experiment was conducted following the Birkbeck College's Ethics Committee. Four participants were excluded due to less than 60% of trials kept due to artefacts in the EEG data. Thus, eighteen participants were included in the analysis (9 females; mean age = 28, SD = 8.09).

Stimuli, design, and Procedure

Experiment one was a basic-level search task in which the target, a picture of a real-world object, was always presented together with 3 distractor objects (see Figure 22).

The participant was instructed to respond as to whether the target appeared in the upper or the lower part of the search array. The target was indicated by a cue that appeared for 200ms, 500ms before the onset of the search array. The cue was either a word indicating which category of objects the target came from or a picture of an exemplar from the target category. Word and Picture Cues were presented in 13 separate blocks (60 trials each), in an ABAB structure with pictorial cue (exact match and picture) and word cues segregated in two different blocks. There were, in fact, three cue conditions, as within the picture cue blocks were randomly presented 60 trials

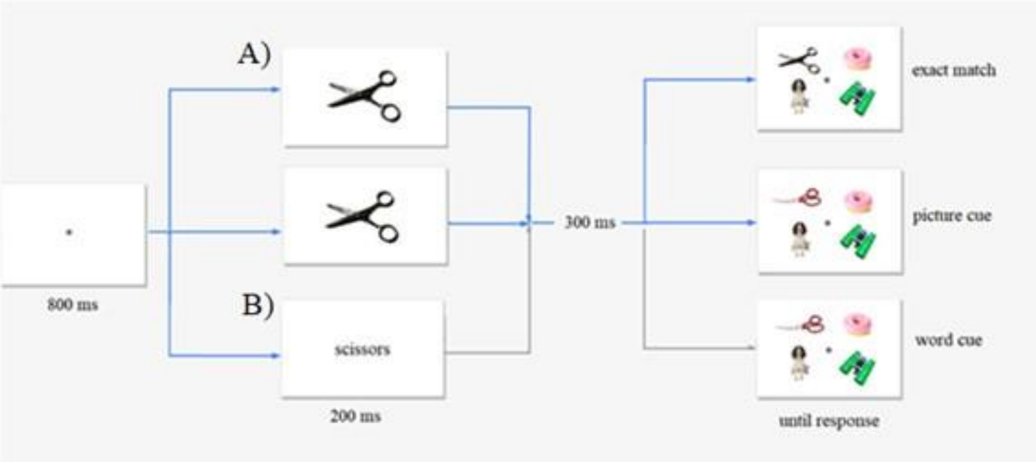


Figure 22: Paradigm of experiment one in which participants performed a visual search task. A fixation cross always preceded the cue by lasting 800 milliseconds. Then, the cue appeared and lasted for 300milliseconds. The three possible cues (exact-match, picture and word) were presented in a blocked design (ABAB) Following the interval intra stimulus (ISI) of 200 milliseconds, the search array always showed the target among three distractors. The search array stayed visible until the participant indicated by pressing a button (arrow up vs arrow down) where the target was presented.

in which the picture cued an exact target, this we refer to as the Exact-match. This design leads to one more block of picture (7) compared to word (6) cues. We selected 30 common object categories with 14 exemplars of each category, from the Object Categories Database (https://cvcl.mit.edu/MM) which provides 401 subcategories of living and non-living things, with 17 examples of each. The stimuli categories were further split into two levels: those categories that had items with a canonical shape (i.e. Basket, Butterfly, Cooking-pot, Donut, Present, Calculator, Guitar, Hat, Jug, Scissors, Sofa, Trousers, Suitcase, Tent and Shoe) and those that did not conform to a canonical shape (i.e. Beer, Binoculars, Buggy, Car, Clock, Dog, Lamp, Leaf, Lock, Phone, Seashell, Speakers, Torch, Vase and Water-Gun) (https://konklab.fas.harvard.edu/#). The cues, the target and the distractor objects were fully randomized within each block and they were presented in a fully random order between participants.

EEG recording and data analysis

EEG was DC-recorded from 23 scalp electrodes at standard positions of the extended 10/20 system (500 Hz sampling rate; 40 Hz low-pass filter) against a left-earlobe reference and was re-referenced offline to averaged earlobes. The continuous EEG was segmented from -100 ms to 500 ms relative to search array onset. Trials with artefacts (horizontal EOG exceeding 30 mV, vertical EOG exceeding 60 mV, all other channels exceeding 80 mV) were removed. Waveforms for trials with correct responses were averaged for each condition separately. N2pc amplitudes were measured at electrodes PO7 and PO8 as ERP mean amplitudes between 190 and 290 ms post-stimulus and N2pc onset defined relative to an amplitude criterion of -1 mV. Participants with less than 50% of trials in any condition (less than 180 correct trials) were removed with the average number of trials per remaining participants being 83%.

Statistical analysis

Both behavioural and physiological data were analysed using Generalized Linear Mixed-effects Models (GLMMs), to overcome the ANOVA assumptions of residuals normality. GLMMs allow to specify the distribution family and the link function that better approximate to parameters of skewed distributions (Ng & Cribbie, 2017) and in addition, they account for both random and fixed effects, with the former being the participants' variability in within-subject designs (Nelder & Baker, 1972). Finally, GLMMs allow to deal with the unequal sample sizes of the Cue Conditions (Exact-match = 60, Picture and Word = 360) (Hesselmann, 2018). For each dependent variable (i.e., reaction times, accuracy and N2pc amplitude) the model that better-approximated data was selected through a model comparison. The likelihood ratio test and the Aikake Information Criterion (AIC) were used as indices of the model’s plausibility, with more plausible models showing significant likelihood ratios and lower AIC.

4.3.2.Results

Behavioural results

Figure 23 shows correct reaction times (RTs >100 ms and < 1000 ms) in all task conditions, separated into Exact-match, Picture and Word, and further split into the Canonical and the Noncanonical subgroups. Since reaction times are distributed with positive skewness and heteroscedasticity, I implemented a GLMM with Gamma distribution and identity link function, considering participants’ variability as a random effect and Cue Condition (Picture, Exact-match and Word) and Shape (Canonical vs Non-canonical) as fixed effects. In Table 8, it is shown the model comparison and selection of the interactive model Cue x Shape.

The interaction showed an obvious differences between the Exact-match (mean: 442ms) and both Word (mean: 548ms, $\beta = 95.46$, $SE = 5.16$, $t = 18.48$) and Picture cues (mean: 530ms, $\beta = 75.8$, $SE = 5.11$, $t = 14.831$), respectively. Moreover, the Picture cue emerged to elicit faster reaction times compared to the Word cue ($\beta = -19.66$, $SE = 3.28$, $t = -5.98$).

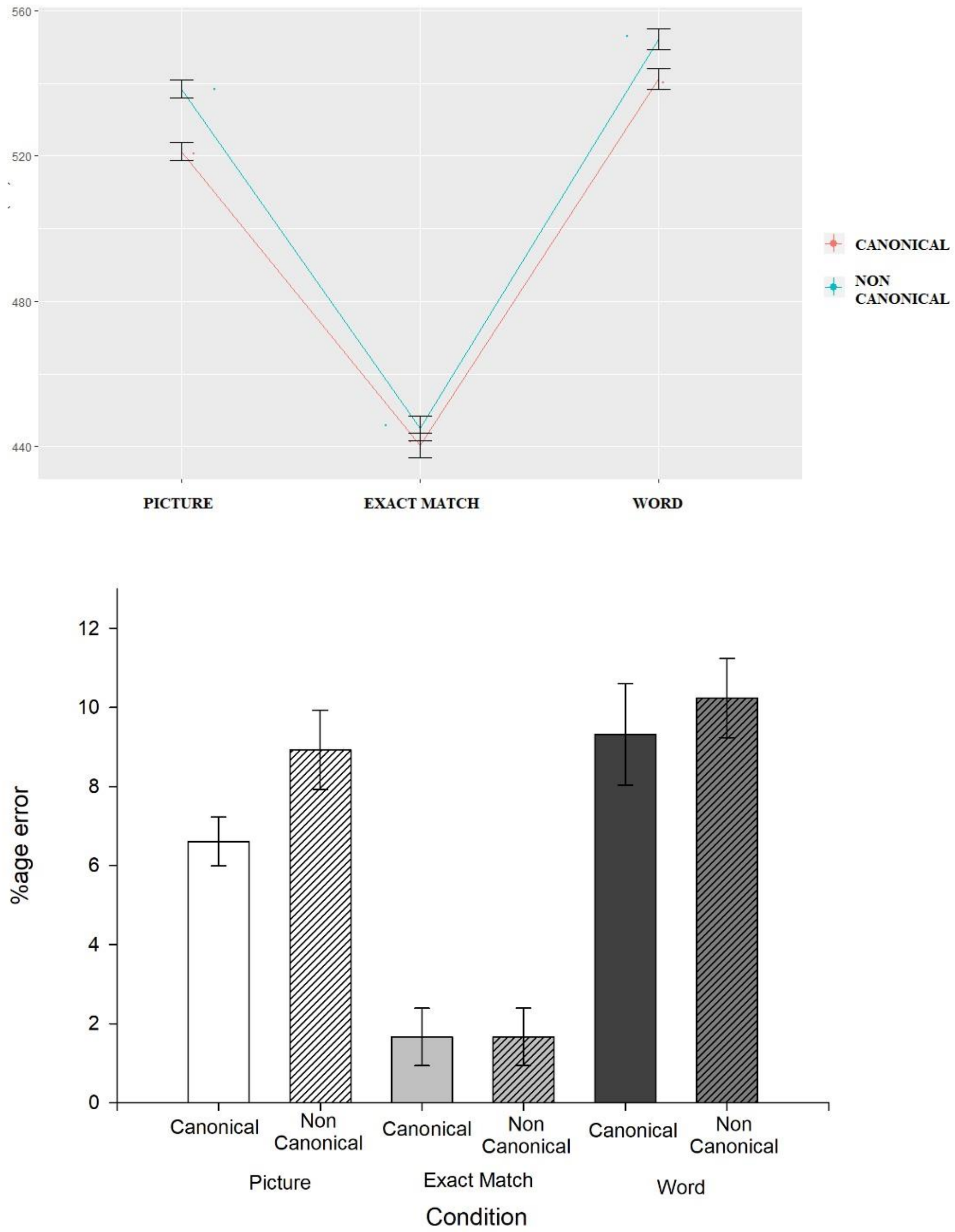
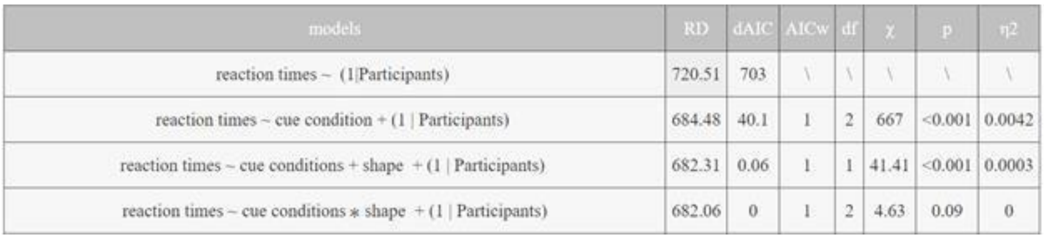


Figure 23. The first plot shows mean correct reaction times (RTs) and the SE and the second plot shows error rate for all trials conditions. Means and error bars are plotted.

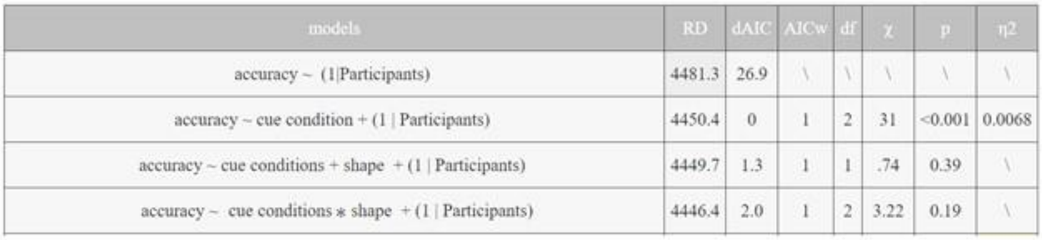


models	RD	dAIC	AICw	df	χ	p	η2
reaction times ~ (1|Participants)	720.51	703	\	\	\	\	\
reaction times ~ cue condition + (1 | Participants)	684.48	40.1	1	2	667	<0.001	0.0042
reaction times ~ cue conditions + shape + (1 | Participants)	682.31	0.06	1	1	41.41	<0.001	0.0003
reaction times ~ cue conditions ∗ shape + (1 | Participants)	682.06	0	1	2	4.63	0.09	0

Table 8. Model comparison for predicting RTs for Cue Condition (Exact-match, Picture and Word) and Shape (Canonical vs Non-canonical). RD = residual deviance, dAIC = difference between a model's Aikake information criterion and those of the best model, AICw = AIC weight, df = degrees of freedom of the chi-squared statistic, η2 = total eta squared as the ratio between the chi-squared and the residual deviance of the null model

The interaction between Shape (Canonical vs Non-Canonical) and Conditions (Exact-match, Picture and Word) showed differences between Canonical and Non-canonical groups both for Picture (Difference of 18ms, $\beta = 17.5$, SE = 3.27, t =5.36) and Word (Difference of 11ms, $\beta = 11.8$, SE = 3.4, t = 3.46), whereas no difference emerged between the Canonical and Non-canonical in the Exact-match condition ($\beta = 3.59$, SE = 6.5, t =0.55).

When the same analysis was performed on accuracy rates (correct trials/ total trials) the results were similar. In this case, I used GLMMs with binomial distribution family and logit link function because accuracy rates residuals are both not normally distributed and a discrete measure. Table 9 shows the model comparison, with a significant main effect for Cue Condition (Exact-match, Picture and Word).



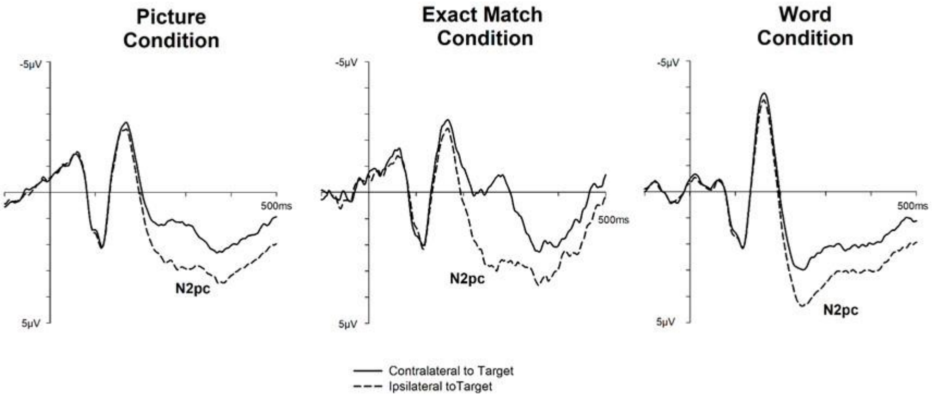
<table>
<tr><th>models</th><th>RD</th><th>dAIC</th><th>AICw</th><th>df</th><th>χ</th><th>p</th><th>η2</th></tr>
<tr><td>accuracy ~ (1|Participants)</td><td>4481.3</td><td>26.9</td><td>\</td><td>\</td><td>\</td><td>\</td><td>\</td></tr>
<tr><td>accuracy ~ cue condition + (1 | Participants)</td><td>4450.4</td><td>0</td><td>1</td><td>2</td><td>31</td><td><0.001</td><td>0.0068</td></tr>
<tr><td>accuracy ~ cue conditions + shape + (1 | Participants)</td><td>4449.7</td><td>1.3</td><td>1</td><td>1</td><td>.74</td><td>0.39</td><td>\</td></tr>
<tr><td>accuracy ~ cue conditions ∗ shape + (1 | Participants)</td><td>4446.4</td><td>2.0</td><td>1</td><td>2</td><td>3.22</td><td>0.19</td><td>\</td></tr>
</table>

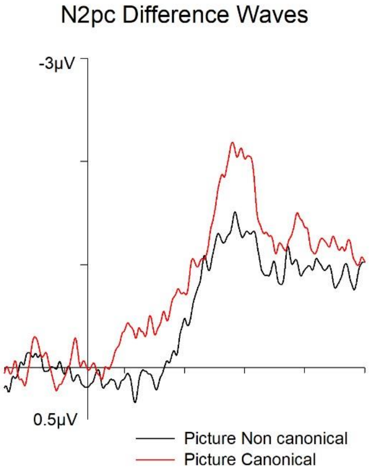
Table 9. Model comparison for predicting Accuracy as the proportion of correct answer on the total trials, for the Cue Condition (Exact-match, Picture and Word) and Shape (Canonical vs Non-canonical). RD = residual deviance, dAIC = difference between a model's Aikake information criterion and those of the best model, AICw = AIC weight, df = degrees of freedom of the chi-squared statistic, η2 = total eta squared as the ratio between the chi-squared and the residual deviance of the null model.

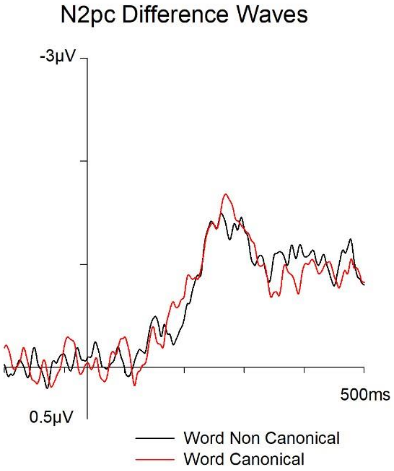
The model comparison (Table 9) did not show any significant main effect for the factor Shape (Canonical vs Non canonical). Whereas the Exact-match condition showed significant higher accuracy rates compared to both the Word (β = -1.26, SE = 0.39, t = -3.23) and Picture cues (β = -1.23, SE = 0.39, t = -3.15). No difference emerged by comparing Word and Picture cues (β = -0.03, SE = 0.12, t = -0.26). Finally, no interaction emerged between Shape and Cue Condition.

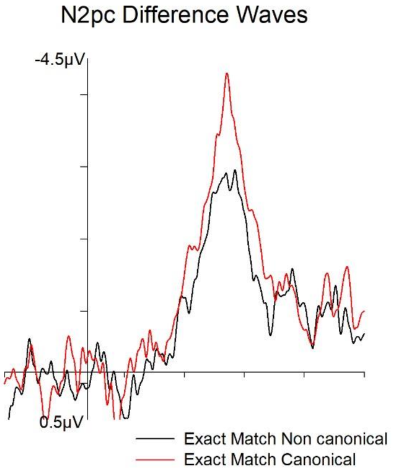
N2pc Component

Figure 24 shows ERPs at electrodes PO7/8 contralateral and ipsilateral to the target in all three cue conditions (Picture, Exact-match and Word) and difference waveforms obtained by subtracting ipsilateral from contralateral ERPs.


Picture Condition
-5µV
500ms
N2pc
5µV
Exact Match Condition
-5µV
500ms
N2pc
5µV
Word Condition
-5µV
500ms
N2pc
5µV
Contralateral to Target
Ipsilateral toTarget


N2pc Difference Waves
-3µV
0.5µV
Picture Non canonical
Picture Canonical


N2pc Difference Waves
-3µV
500ms
0.5µV
Word Non Canonical
Word Canonical


N2pc Difference Waves
-4.5µV
0.5µV
Exact Match Non canonical
Exact Match Canonical

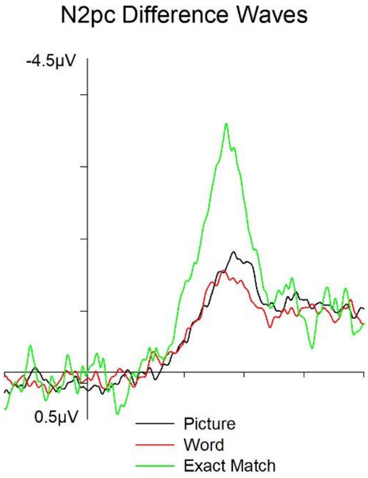

N2pc Difference Waves
-4.5µV
0.5µV
Picture
Word
Exact Match

Figure 24. In the upper side: ERPs at electrodes PO7/8 contralateral and ipsilateral to the target in all three cue conditions Picture, Exact-match and Word, and difference waveforms obtained by subtracting ipsilateral from contralateral ERPs. In the downside**:** Difference waveforms for all the cue conditions Picture and Word separated into Canonical and Non-canonical.

Firstly, the comparison between contralateral and ipsilateral ERP mean amplitudes confirmed the presence of N2pc components locked to target onset and cued by Word (β = 0.16, SE = 0.05, t = 2.96), Picture (β = 0.15, SE = 0.06, t =2.5) and Exactmatch (β = -0.16, SE = 0.06, t =-2.96).

In table 10, the model comparison predicting ERP’s average amplitudes in the 190-290 ms post-stimulus, shows the significant main effects of Cue Condition and Laterality (electrode contralateral versus ipsilateral to the target), whereas it did not emerge any main effect of Shape. Furthermore, the interactive model, including the three ways interaction among Shape, Laterality and Cue Condition, was selected as the best model.

models	RD	dAIC	AICw	df	χ	p	η2
Amplitude [mV] ~ (1|Participants)	956	56.3	\	\	\	\	\
Amplitude [mV] ~ shape (canonical vs. non-canonical) + (1 | Participants)	918.2	24.3	1	1	.28	0.59	\
Amplitude [mV]~ laterality (left vs. right) + (1 | Participants)	955.3	58.9	1	0	.14	<0.001	0.0002
Amplitude [mV]~ cue condition (Picture, Word, Eact match) + (1 | Participants)	955.2	58.7	1	1	37	<0.001	0.3
Amplitude [mV] ~ cue condition * shape * laterality + (1 | Participants)	880.5	0	1	8	37.6	<0.001	0.3

Table 10. Model comparison for predicting Amplitude N2pc for the Cue Condition (Exact-match, Picture and Word) and Shape (Canonical vs Non-canonical). RD = residual deviance, dAIC = difference between a model's Aikake information criterion and those of the best model, AICw = AIC weight, df = degrees of freedom of the chi-squared statistic, η2 = total eta squared as

the ratio between the chi-squared of each model and the residual deviance of the null model.

The three-way interaction shows that Canonical and Non canonical Shape did not differ in the Word cue (β = 0.03, SE = 0.03, t = 1.16) and in the Exact-match cue (β = -0.20, SE = 0.60, t = -0.33) whereas they showed a significant difference in the Picture cue (β = -1.05, SE =0.29, t =-3.57). In the Cue Condition, the N2pc exhibited higher negative amplitude in the Exact-match with respect to both the Word (β = -1.53, SE = 0.21, t = -7.04) and the Picture cues (β = 0.25, SE = 0.04, t = 5.45). No difference emerged by comparing the Word and the Picture cue (β = -0.005, SE = 0.25, t = -0.02).

4.3.3. Discussion experiment 1

During category search among basic-level categories, the picture cues (e.g., the picture of a pen to search for a pen) did trigger faster search time compared to the word cues (e.g., the word pen to search for a pen) (Wolfe, Horowitz, Kenner, Hyle, & Vasan, 2004a). However, the two cues did not differ in accuracy rates during target selection, neither they show difference during target selection as shown by N2pc components analysis.

Furthermore, canonical categories emerged to play a significant role during the category-guided search by prompting faster reaction times compared to Noncanonical categories, for both picture and word cues. Accordingly, the analysis of N2pc components showed higher amplitudes for the Canonical categories compared to Non-canonical categories.

However, the effect emerged to be significant only with the picture cue by suggesting that category-guided search cued by words was less sensitive to the perceptual distinctiveness of category-items. The RTs findings of experiment one can be described in terms of an advantage of pictorial cues in guiding category search and corroborated the idea that the category template triggered by visual cues is influenced by canonical representations (Lupyan & Spivey, 2010; Schmidt & Zelinsky, 2009b). Altogether, the results suggest that the target representation cued by pictorial rather than conceptual information, facilitate category-guided search.

Experiment two was set to disentangle those interpretations, which may be easily explained by a visual perceptual advantage prompted by both, the canonical (high perceptual distinctiveness) and the basic-level (high conceptual distinctiveness) categories (Konkle et al., 2010), interacting in experiment one. Indeed, the high distinctiveness of basic-level categories (e.g., a pen to search for a pen) might not require participants to categorise the target at a semantic level when pictures cued it. This manipulation might unbalance the saliency of the visual features in WM at the expense of the conceptual features. Thus, the strategy of relying on visual similarity rather than on conceptual similarity may be a preferential and successful strategy during such category-guided search.

In experiment two, I investigated if these results were confirmed at a higher level of the hierarchy of category classification (e.g., stationery). Experiment two focused on templated-guided search for supraordinate category-items (e.g., stationery to look for a pen or a notebook) whose items show lower perceptual and conceptual distinctiveness compared to basic-level categories (Konkle et al., 2010; Maxfield & Zelinsky, 2012; E. H. Rosch, 1973). I aimed to shed light on the nature (perceptual or conceptual) of mental representation triggered by either picture and word cues further interacting with Canonical and Non-canonical categories, during such categorical search. If the template-guided search for a category item is more likely to be conceptual, I expected to find an advantage for supraordinate category word cues compared to basic-level category and picture cues. Indeed, superordinate categories are necessarily conceptual by pointing to wider groups of items in the category hierarchy (e.g., stationery). However, supraordinate category cues further interacting with dense perceptual categories (canonical categories) should indicate that the visual features of the target are still represented in a canonical fashion, during category-guided search.

4.3.4. Experiment 2: Method

Participants

Twenty-one English native speaking participants with normal or corrected to normal vision (11 females; mean age = 28 years old, SD = 9.06 years) took part in the experiment two and were reimbursed for their participation. All participants signed informed consent, and the experiment was conducted following the Birkbeck College's Ethics Committee. Five participants were excluded due to less than 60% of trials kept due to artefacts in the EEG data. Thus, eighteen participants were included in the analysis (9 females; mean age = 28, SD = 9.43).

Stimuli, design, and Procedure

Experiment 2 was a basic category search task in which the target, a real-world object, was always presented among 3 distractor objects and the participant was instructed to respond as to whether the item belonging to the same category pointed by the cue appeared in the upper or lower visual field. See Figure 25. The target was indicated by a cue that appeared for 200ms, 500ms prior to the onset of the search array. The cue was either a word indicating which basic-level (e.g., pen) or supraordinate (e.g., stationery) category of objects the target came from or a picture of an exemplar from the target category. Word and Picture Cues were presented in separate blocks, in an ACBACB structure (see Figure 25). There were, in fact, three cue conditions (i.e., basic-level category, supra-ordinate category and picture), each of 240 trials (720 total trials). I selected 16 common object categories with 15 exemplars of each category, from the Object Categories Database (https://cvcl.mit.edu/MM) as candidate category targets. In addition, I selected other 16 common object categories with 15 exemplars of each category being randomly presented as distractor objects in the search array. The stimuli categories were further split into two levels, constituted by 8 categories whose items shared low perceptual variance namely, Canonical (i.e., animal, body part, clothing, drink, food, insect, vehicle) and by 8 categories whose items showed high perceptual variance namely, Non-canonical (i.e., building, furniture, game, jewellery, kitchen item, medical equipment, stationery, weapon, music instrument, Konkle et al.; 2010). Being constrained by this structure, the cues, the targets and the distractor objects were fully randomised within 18 blocks (40 trials each) and between both conditions and participants.

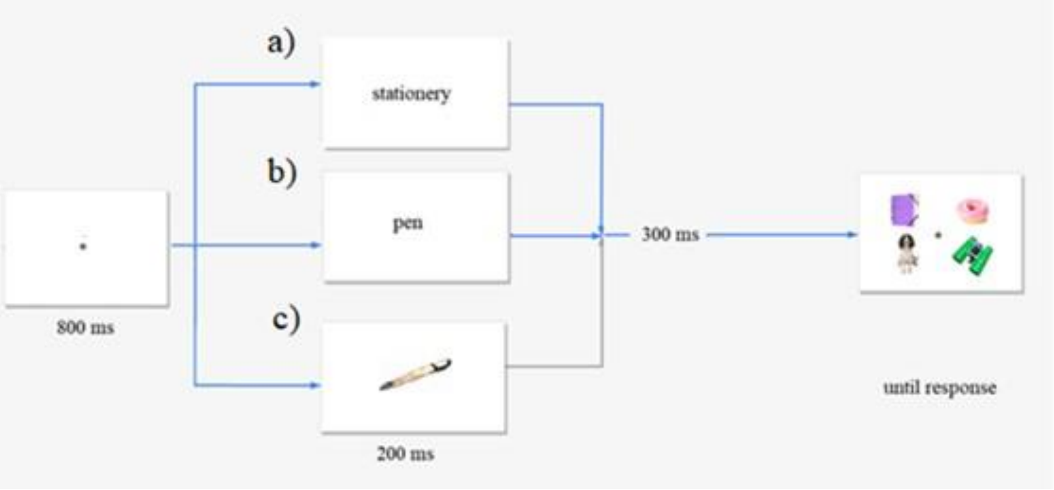


Figure 25. Paradigm of experiment 2 in which participants performed a categorical search task. A fixation cross always preceded the cue by lasting 800 milliseconds. Then, the cue appeared and lasted for 200milliseconds. The three possible cues a) supraordinate category, b) basic-level category and c) picture were presented in a blocked design (ACBACB). Following an (ISI) of 300 milliseconds, the search array always showed the target among three distractor objects. The search array stayed visible until participant indicated the target position by pressing a button (arrow up vs arrow down).

4.3.5. Results

Behavioural results

RTs (> 100 ms and < 1000 ms) data were modelled using GLMMs with Gamma distribution and identity link function, considering participants' variability as a random effect and Cue Condition (Supra-ordinate category, Basic-level category and Picture) and Shape (Canonical vs Non-canonical) as fixed effects. See Figure 26.

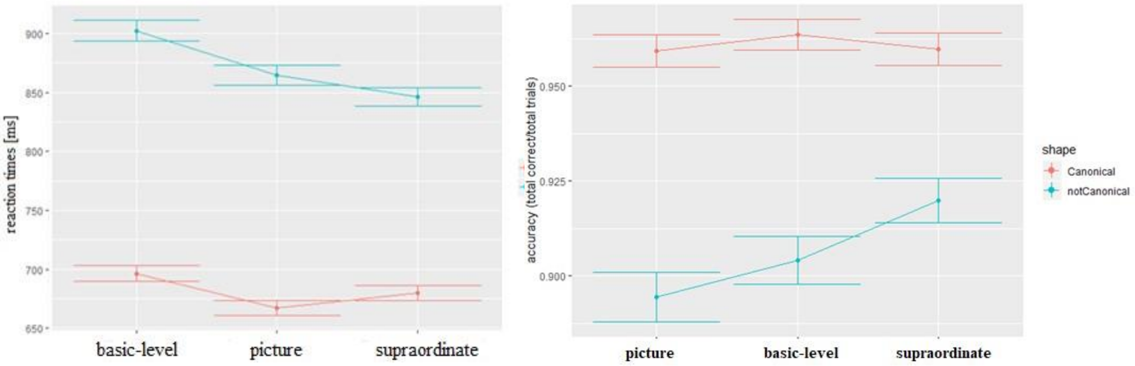


Figure 26. Interaction plots of correct RTs (on the left panel) and accuracy rate (correct trials/ total trials) on the right panel for Cue Conditions (picture, basic level and supraordinate level) further split for Canonical (red) and Non-canonical (green) shape of category-items. Standard errors are plotted.

The model comparison showed the interactive model Cue Condition X Shape to better explain data (dAIC = 0, $\chi2(1)$ = 1221.2, p < 0.001).

The Shape showed significantly faster RTs for Canonical compared to Non canonical categories both for Picture (β = 186.14, SE = 9.97, t = 18.67), Basic-level category (beta = 199.04, SE = 10.08,t = 19.74) and Supra-ordinate category word cues (β = 158.2, SE = 9.15, t = 17.30). Finally, the interaction revealed significant longer RTs for Non canonical categories cued by Basic-level categories compared to both Picture (β = -29.91, SE = 7.12, t = -4.20,) and Supraordinate cues (β = 30.51, SE = 7.09, t =-4.30), no significant difference emerged between the Picture and the Supraordinate cues (β = 0.60, SE = 6.93, t =0.09).

The same analysis was performed on accuracy rates, using GLMMs with binomial distribution family and logit link function because accuracy rates residuals are both not normally distributed and a discrete measure. The model comparison showe the interactive model Cue Condition x Shape to better approximate data (dAIC = 0, $\chi2(1)$ = 170, p < 0.001). The Canonical categories triggered faster RTs compared to Non canonical categories for Picture (β = -1.05, SE = 0.13, t = -8.02), Basic-level category (β = -1.07, SE = 0.14, t =-7.71) and Supraordinate category cues (β = -0.76, SE = 0.14, t = -5.51),

Furthermore, the interaction shows higher accuracy for Non canonical Supraordinate categories compared to Non canonical Picture cues (β = -0.22, SE = 0.09, t = -2.46) whereas not difference emerged with respect to Non canonical Basic-level category cue (β = -0.10, SE = 0.09, t = -1.17). Finally, the comparison between Non canonical Basic-level category and Picture cues did not lead to any significant difference (β = 0.11, SE = 0.09, t = 1.29).

N2pc Components

Figure 27 shows the difference waveforms of grand averaged N2pc amplitude for all the Cue Conditions (Picture, Basic-level and Supraordinate) separated into Canonical and Noncanonical.

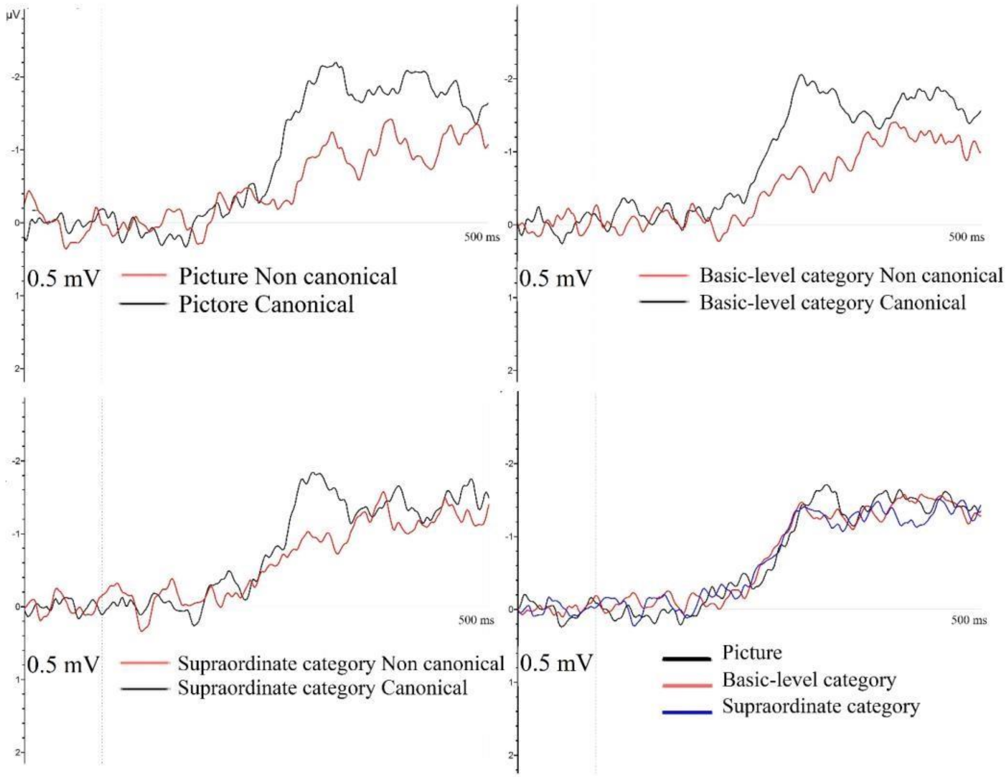


Figure 27. Difference wave Difference waveforms of grand averaged N2pc amplitude for all the Cue Conditions (Picture, Basic-level and Supraordinate) separated into Canonical and Non-canonical.

The three-way interactive model among Shape, Laterality and Cue Condition emerged to be the best model (dAIC =14, $\chi2(1) = 32.09$, $p < 0.001$).

With regards to Cue Condition, the N2pc did not show any difference in amplitude between the Supraordinate category and both the Basic-level category (β = 0.03, SE = 0.16, t =

0.22) and the Picture cue (β = 0.05, SE = 0.16, t = 0.37), as shown in figure 8. No difference emerged neither by comparing the Basic-level category and the Picture cues (β = -0.02, SE = 0.17, t = -0.13).

The interaction showed that Canonical categories triggered higher N2pc amplitude compared to Non canonical categories (β = 0.65, SE = 0.11, t = 5.92) in all Supraordinate category (β = 0.5, SE = 0.27, t = 3.36), Basic-level category (β = 0.37, SE = 0.27, t = 1.36) and Picture cues (β = -1.05, SE = 0.29, t = -3.57), as shown in Figure 28.

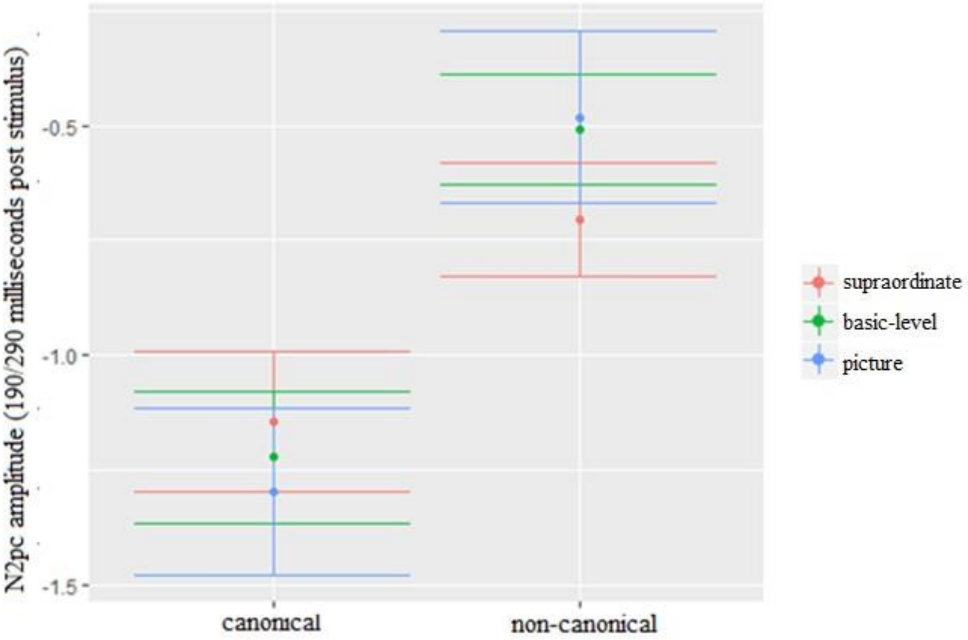


Figure 28. Interaction plot of the averaged N2pc amplitude 190/290 milliseconds post-stimulus for the Canonical and Non-canonical shape of category-items further split for Cue Conditions (supra-ordinate category, basic-level category and picture). Standard errors for repeated measures are plotted.

4.3.6. Discussion experiment 2

The findings of experiments two confirmed and extended those of experiment one. In fact, the pictorial-guided search (e.g. a pen to search for stationery items) was significantly faster compared to the basic-level word cue (e.g., pen to search for stationery items). In addition, the supraordinate word cues (e.g., stationery) did share the same advantage of picture cues. Moreover, participants showed a significantly higher accuracy in the supraordinate category-guided search compared to both the basic-level word and the picture cues, while searching for category-items with low perceptual distinctiveness (noncanonical shape). Nevertheless, categories with high perceptual distinctiveness (canonical shape) significantly promoted a faster and more accurate target selection compared to noncanonical categories independently from the cue condition, and N2pc amplitude analysis showed that during the early stages of perceptual selection canonical shape trigger target selection compared to noncanonical shapes. Moreover, N2pc did not show any difference across Cue Conditions. Then, these findings suggest that during category-guided search among items with low perceptual and conceptual distinctiveness, the category template relies on linguistic-mediated (category classification) and the perceptual (canonical and noncanonical shape) representations, rather than relying uniquely on perceptual representations (Goldstone, 1994b, 1994a; Lupyan, 2015; Perry & Lupyan, 2017).

4.3.7. General discussion

In human language, nouns typically convey categorical information and always point to more than one specific object. For instance, the content word *pen* can be used to refer to many different entities in the real world. In fact, even if it points to a basic-level category, the target – which exact pen the cue is referring to – stays unspecified (Edmiston & Lupyan, 2015). In addition, the number of different objects pointed by a single word rise as a function of the conceptual hierarchy prompted by the context (Lupyan, Thompson-schill, & Swingley, 2010; Rosch, 1973). In two experiments, I explored the conditions under which pictorial and word cues produce equivalent guidance and the formation of categorical target representations. The Label Feedback Hypothesis (Lupyan, 2007, 2012) expects words to cue a more abstract mental representation of the target; hence, words should promote a more efficient category-guided search compared to picture cues. Evidence show that word cues successfully guide the selection of category items (Wolfe, Horowitz, Kenner, Hyle, & Vasan, 2004b) and do elicit the N2pc component (Nako, Wu, Smith J., et al., 2014). However, visual search literature suggests that target guidance should always be better with a pictorial rather than a categorical cue (Nako, Wu, & Eimer, 2013; Nako, Wu, Smith, & Eimer, 2014; Schmidt & Zelinsky, 2009b; Yang & Zelinsky, 2009).

In experiment one, participants were cued to category-items of basic-level categories (e.g., pen to search for a pen). In experiment two, participants were cued to search for items of supraordinate categories (e.g., stationery or pen to search for a notebook). Then, the manipulation of the hierarchy level along the category classification prompted by the context (e.g., pen or stationery) was expected to show word cues triggering the category item selection (Lupyan, 2019; Lupyan & Ward, 2013). Thus, I expected to find a word facilitation effect in terms of accuracy, RTs, and higher N2pc amplitude as a function of the conceptual hierarchy prompted by the context of each experiment. These expectations were partially confirmed. In experiment one, the type (i.e., word or picture) and specificity (i.e., canonical and noncanonical) of the cues used in the visual search task influenced how much information could be loaded into the search template, which in turn influenced participants' performance (Wolfe et al., 2004b). The results showed that picture cues led to a visual superiority effect. However, this effect might reflect a basic-level advantage prompted by the categorical classification of experiment one. In fact, basic-level items are perceptually similar in appearance (e.g., pens) and in turn, they are more rapidly categorized at the perceptual level (Murphy & Brownell, 1985; E. Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976; E. Rosch, Simpson, & Miller, 1976) compared to supraordinate category-items. Categorisation theories typically explain the basic-level advantage in terms of a favourable balance existing between perceptual and conceptual distinctiveness for basic-level categories (Murphy & Brownell, 1985). Shape information at the basic level also cue the category membership (Maxfield & Zelinsky, 2012; E. Rosch, Simpson, et al., 1976; Wolfe & Horowitz, 2004), potentially adding an advantage to picture cues in such category-guided search tasks. Hence, in experiment one, the same item was recognised as a pen (basic level category) more rapidly with picture cues compared to word cues, because of its perceptual distinctiveness. Another explanation not mutually exclusive with the previous is that the word disadvantage found in experiment one might be explained by the fact that representations of basic-level categories are more useful for target verification than for guidance, as basic representations typically lack the specificity that emerged to be relevant for guided-search. These interpretations are in fact, in line with visual search literature suggesting that target guidance should always be better with a pictorial rather than a word cue (Schmidt & Zelinsky, 2009a; Wolfe & Horowitz, 2004; Wolfe et al., 2004b). However, supraordinate categories are necessarily conceptual and point to a wide group of items from higher hierarchy level e.g., stationery. Similarly, the features of superordinate items are conceptually distinct, but the variability at the perceptual level means that the features of search template generally lack specificity. Then, the LFH expects no pictures to be as much efficient as the label stationery in cueing the representation of the target, during search for anything to take a note. This seemed to be the case of experiment two. Our failure to replicate a pictorial advantage during category-guided search for supraordinate items means one of two things: either the mental representation of the target cued by supra-ordinate words indicated a facilitatory effect of word or the target representation cued by pictures did not guided search as efficiently as expected, and in this case no better than word cues. The main results of this study corroborate what is expected by the reverse hierarchy theory of information processing (Reverse Hierarchy Theory RHT) (Ahissar & Hochstein, 2004), in which the perceptual categorisation is defined as the interplay between visual and conceptual functions. In fact, the supraordinate category-guided search further interacted with high perceptual distinct categories (canonical categories) by showing to be influenced also by perceptual representation. To conclude, these findings indicated that categorical-attention mechanisms are influenced both by conceptual and perceptual distinctiveness of category-items. Future works may aim to disentangle the role played by subordinate categories and word order (e.g. red pen and pen) that are expected to trigger further the basic-level advantage compared to picture cues, during category-based search (Freedman, 2001; Zelinsky et al.; 2008). Furthermore, it would be useful to investigate the degree at which the perceptual categorisation of novel objects (not influenced by individual differences in terms of familiarity) is affected by the interaction of nonsense linguistic concomitant associations and shape variation during active search for category-items.

4.3.8. Bibliography

Ahissar, M., & Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. Trends in Cognitive Sciences, 8(10), 457–464.

Blair, M. R., Watson, M. R., & Meier, K. M. (2009). Errors, efficiency, and the interplay between attention and category learning. Cognition, 112(2), 330–336.

Clarke, A. (2017). Top-down cognitive influences on object recognition: a commentary on Perry and Lupyan (2016). Language, Cognition and Neuroscience, 32(8), 947–949.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. Annual Review of Neuroscience, 18(1), 193–222.

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. Psychological Review, 96(3), 433–458. https://doi.org/10.1037/0033-295X.96.3.433

Edmiston, P., & Lupyan, G. (2015). What makes words special ? Words as unmotivated cues. Cognition, 143, 93–100. https://doi.org/10.1016/j.cognition.2015.06.008

Eimer, M. (1996). The N2pc component as an indicator of attentional selectivity. Electroencephalography and Clinical Neurophysiology, 99(3), 225–234.

Eimer, M. (2014). The neural basis of attentional control in visual search. Trends in Cognitive Sciences, 18(10), 526–535. https://doi.org/10.1016/j.tics.2014.05.005

Eimer, M., & Grubert, A. (2014). The gradual emergence of spatially selective target processing in visual search: From feature-specific to object-based attentional control. Journal of Experimental Psychology: Human Perception and Performance, 40(5), 1819–1831. https://doi.org/10.1037/a0037387

Freedman, D. J. (2001). Categorical Representation of Visual Stimuli in the Primate Prefrontal Cortex. 291(January).

Goldstone, R. L. (1994a). Influences of categorization on perceptual discrimination. Journal of Experimental Psychology: General, 123(2), 178.

Goldstone, R. L. (1994b). The role of similarity in categorization: Providing a groundwork. Cognition, 52(2), 125–157.

Goldstone, R. L., Kersten, A., & Carvalho, P. F. (2018). Categorization and concepts. Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience, 3, 1–43.

Harnad, S. (1987). Psychophysical and cognitive aspects of categorical perception: A critical overview. In Categorical perception: The groundwork of cognition (pp. 1– 52). Cambridge University Press.

Hesselmann, G. (2018). Applying Linear Mixed Effects Models (LMMs) in WithinParticipant Designs With Subjective Trial-Based Assessments of Awareness — a Caveat. 9(May), 1–5. https://doi.org/10.3389/fpsyg.2018.00788

Jenkins, M., Grubert, A., & Eimer, M. (2018). Category-based attentional guidance can operate in parallel for multiple target objects. Biological Psychology, 135(November 2017), 211–219. https://doi.org/10.1016/j.biopsycho.2018.04.006

Josephs, E., & Konkle, T. (2017). Object, scenes, and the spaces in between: Workspaces have distinctive perceptual and semantic content. Journal of Vision, 17(10), 1093.

Konkle, T., Brady, T. F., & Alvarez, G. A. (2010). Conceptual Distinctiveness Supports Detailed Visual Long-Term Memory for Real-World Objects. 139(3), 558–578. https://doi.org/10.1037/a0019165

Konkle, T., & Oliva, A. (2012). A real-world size organization of object responses in occipitotemporal cortex. Neuron, 74(6), 1114–1124.

Luck, S. J., & Hillyard, S. A. (n.d.). Spatial filtering during visual search: Evidence from human electrophysiology. Journal of Experimental Psychology: Human Perception and Performance, 20(5), 1000–1014. https://doi.org/10.1037/0096-1523.20.5.1000

Lupyan, G. (2012). Linguistically modulated perception and cognition: The labelfeedback hypothesis. Frontiers in Psychology, 3(MAR), 1–13. https://doi.org/10.3389/fpsyg.2012.00054

Lupyan, G. (2015). How Language Programs the Mind. 1–17. https://doi.org/10.1111/tops.12155

Lupyan, G. (2019). Target representation during category-based visual search : is there any word advantage ? 1(January), 2012–2014. https://doi.org/10.1037/xhp0000343

Lupyan, G., & Spivey, M. J. (2010). Making the Invisible Visible : Verbal but Not Visual Cues Enhance Visual Detection. 5(7), 1–9. https://doi.org/10.1371/journal.pone.0011452

Lupyan, G., Thompson-schill, S. L., & Swingley, D. (2010). Psychological Science. (March). https://doi.org/10.1177/0956797610366099

Lupyan, G., & Ward, E. J. (2013). Language can boost otherwise unseen objects into visual awareness. https://doi.org/10.1073/pnas.1303312110/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1303312110

Maxfield, J. T., Stalder, W. D., Brook, S., & Zelinsky, G. J. (2014). Effects of target typicality on categorical search. 14, 1–11. https://doi.org/10.1167/14.12.1.doi

Maxfield, J. T., & Zelinsky, G. J. (2012). Searching through the hierarchy: How level of target categorization affects visual search. Visual Cognition, 20(10), 1153–1163.

Mazza, V., Turatto, M., & Caramazza, A. (2009). Attention selection, distractor suppression and N2pc. Cortex, 45(7), 879–890.

Murphy, G. L., & Brownell, H. H. (1985). Category differentiation in object recognition: typicality constraints on the basic category advantage. Journal of Experimental Psychology: Learning, Memory, and Cognition, 11(1), 70.

Nako, R., Smith, T. J., & Eimer, M. (2004). The Role of Color in Search Templates for Real-world Target Objects. 1714–1727. https://doi.org/10.1162/jocn

Nako, R., Wu, R., & Eimer, M. (2013). Rapid Guidance of Visual Search by Object Categories. 39(4), 1–11. https://doi.org/10.1037/a0033228

Nako, R., Wu, R., Smith J., T. J., & Eimer, M. (2014). Item and category-based attentional control during search for real-world objects: Can you find the pants among the pans? Journal of Experimental Psychology: Human Perception and Performance, 40(4), 1283–1288. https://doi.org/10.1037/a0036885

Nako, R., Wu, R., Smith, T. J., & Eimer, M. (2014). Item and Category-Based Attentional Control During Search for Real-World Objects : Can You Find the Pants Among the Pans ? 40(4), 1283–1288.

Nelder, J. A., & Baker, R. J. (1972). Generalized linear models. Wiley Online Library.

Ng, V. K. Y., & Cribbie, R. A. (2017). Using the gamma generalized linear model for modeling continuous, skewed and heteroscedastic outcomes in psychology. Current Psychology, 36(2), 225–235.

Olivers, C. N. L., Peters, J., Houtkamp, R., & Roelfsema, P. R. (2011). Different states in visual working memory: When it guides attention and when it does not. Trends in Cognitive Sciences, 15(7), 327–334.

Perry, L. K., & Lupyan, G. (2017). Recognising a zebra from its stripes and the stripes from " zebra ": the role of verbal labels in selecting category relevant information. 3798. https://doi.org/10.1080/23273798.2016.1154974

Rosch, E. H. (1973). On the internal structure of perceptual and semantic categories. In Cognitive development and acquisition of language (pp. 111–144). Elsevier.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. Cognitive Psychology, 8(3), 382–439.

Rosch, E., Simpson, C., & Miller, R. S. (1976). Structural bases of typicality effects. Journal of Experimental Psychology: Human Perception and Performance, 2(4), 491.

Schmidt, J., & Zelinsky, G. J. (2009a). Search guidance is proportional to the categorical specificity of a target cue. The Quarterly Journal of Experimental Psychology, 62(10), 1904–1914.

Schmidt, J., & Zelinsky, G. J. (2009b). Short article Search guidance is proportional to the categorical specificity of a target cue. 62(10), 1904–1914. https://doi.org/10.1080/17470210902853530

Schneider, D., Beste, C., & Wascher, E. (2012). On the time course of bottom-up and top-down processes in selective visual attention: An EEG study. Psychophysiology, 49(11), 1660–1671.

Wolfe, J. M., & Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it? 5(June), 1–7.

Wolfe, J. M., Horowitz, T. S., Kenner, N., Hyle, M., & Vasan, N. (2004a). How fast can you change your mind ? The speed of top-down guidance in visual search. 44, 1411–1426. https://doi.org/10.1016/j.visres.2003.11.024

Wolfe, J. M., Horowitz, T. S., Kenner, N., Hyle, M., & Vasan, N. (2004b). How fast can you change your mind? The speed of top-down guidance in visual search. Vision Research, 44(12), 1411–1426. https://doi.org/10.1016/j.visres.2003.11.024

Yang, H., & Zelinsky, G. J. (2009). Visual search is guided to categorically-defined targets. Vision Research, 49(16), 2095–2103.

4.4. Into the time course of sentence verification and visual target selection[7]

The meaning of the sentence *Mary is not guilty* (Mary is innocent) is easier to comprehend as compared for instance, to the sentence *Mary is not French* (Mary could be either American, Italian etc). By means of contradictory predicates (e.g. active/passive, not open/close) the reader requires no extra steps to understand that *not guilty* means innocent (Mayo, Schul, & Burnstein, 2004; Hasson & Glucksberg, 2006; Kaup, Lüdtke, & Zwaan, 2006a). Thus, the special case of contradictory predicates shows that the named and the actual state of affairs conveyed by negative and affirmative sentences can prompt a symmetrical truth value. Moreover, it shows that negation can be used to assert or deny depending on the context of use (Giora, Fein, & Aschkenazi, 2006). Nevertheless, some pragmatic inferences are necessary to understand why “not guilty” has been used instead of “innocent”.

---

[7] This study has been carried out in collaboration with Francesco Vespignani, Max Paulus and Veronica Mazza. Data were collected at the EEG Lab, Centre Interdepartmental of Mind and Brain, University of Trento.

The present study investigated the verification of contradictory predicates with the aim to disambiguate the interpretation of affirmative and negative sentence when there is a unique interpretation of the sentence (e.g., Mary is not guilty = Mary is innocent) compared to when there are multiple possible interpretations of the sentence (Mary is not French = Mary could be either English, Italian; Spychalska, Haase, Kontinen, & Werning, 2016). Importantly, here contradictory predicates refer to adjectives of visual objects like the colour, i.e., black/white. On note, even though any grey or coloured figure would match with not black in this specific set context by presenting only white and black figures the unique interpretation of not black points to white. This should further favour the likelihood of computing the actual state of affairs during sentence comprehension of negative sentences (Tian, Breheny, & Ferguson, 2010b).

Research on sentence comprehension has shown longer processing times and higher error rates for negative compared to affirmative sentences verification (Carpenter & Just, 1975; Haviland & Clark, 1974; Kaup, Lüdtke, & Zwaan, 2006b; Kaup & Zwaan, 2003). In linguistics, this main finding has been explained by the traditional argument that negative sentences are in fact affirmative sentences with a negative element; thus, one more element to elaborate. Thus, the longer times observed should reflect the process of the proposition itself and its negation (Jordan, 1998; Tettamanti et al., 2008).

At a more representational level, Kaup et al., (2006) proposed the two-steps simulation hypothesis (TSSH) in order to explain the classical disadvantage shown by negative with respect to affirmative sentence verification. This hypothesis predicts a special status for mental representation of negative sentences by embracing a non-incremental perspective. That is, in order to get the meaning of a sentence like *The door is not open* individuals first represent the negated state of affair ('an open door') and then the actual state of affair ('a closed-door'). Several authors have challenged this hypothesis. For instance, Tian et al. (2016, 2010) investigated cleft and noncleft sentences by manipulating their polarity. The results showed that the simulation of the negated state of affair prior to the one of the actual state of affair it is not a mandatory step (Tian & Breheny, 2007; Tian, Ferguson, & Breheny, 2016). In fact, it seems more likely that the actual state of affair may be represented directly depending on the sentence structure and the situated context, according to a more dynamic account of sentence processing. Finally, somewhere in the middle between propositional theories and perceptual simulation, we find the Mental Model theory (Johnson-Laird, 1983). This theory predicts that mental representations used during comprehension, are simulations (mental models) that do not directly represent the linguistic input instead they simulate the relationship between the elements in the sentence, through intermediate representations. These intermediate representations are logical operators (e.g., not, if, for) applied by individuals on sentence elements to figure the actual state of affairs (Hasson & Glucksberg, 2006). The Mental Model theory expects individuals to understand negation by simulating all the possible representations cued by a sentence under the knowledge individuals have about the context (Khemlani, Orenes, & Johnson-Laird, 2012) and depending on individuals working-memory capacity (concomitant representations activated; Ort, Fahrenfort, Ten Cate, Eimer, & Olivers, 2019).

Among other paradigms, the Sentence Picture Verification (SPV) task has been widely used to grasp insights on the role of perceptual representation triggered by sentences (Just & Carpenter, 1971). In fact, by presenting a sentence prior of a related picture and by asking participants to verify the truth value of the sentence, this task allows to measure the degree by which the representation prompted by linguistic information and the pictorial information do match. However, the debate about the role of mental representation triggered by negative sentences is still open, and the evidence that supports each theory are somehow inconsistent.

From the side of the picture verification, early EEG studies have shown that when individuals are asked to search for a picture, they rely on its perceptual representation to further select the right target (Eimer, 1996). Moreover, consistent evidence showed that the less the perceptual representation (triggered by pictorial cues) and the target match the more the correct picture selection seems to be disrupted (Nako, Wu, Smith, & Eimer, 2014). From a semantic point of view, the potential need to suppress positive information is also likely to increase processing costs (Spychalska, Haase, Kontinen, & Werning, 2016; Giora, Heruti, Metuki, & Fein, 2009). Accordingly, in visual search task target-present trials do trigger picture selection compared to target-absent trials (Pashler, 1987). Thus, it seems that during picture verification, the cognitive system is template-guided by the pictorial representation of the target and importantly, this perceptual template seems to be recruited also during sentence-guided search. For instance, the TSSH expects that the representation triggered by either affirmative or negative sentences should rely on perceptual simulations (Kaup et al., 2006a). For example, it has been shown that when the picture of a flying eagle follow a sentence like *The eagle is in the sky* participants are faster in judging the truth value of this sentence, compared to a sentence like *The eagle is in the nest* preceding the picture of a flying eagle (Kaup et al., 2006b; Tian et al., 2010b). In an EEG study by Lüdtke, Friedrich, De Filippis, & Kaup, (2008b), the authors presented negative and affirmative sentence cues, word by word to participants that were asked to evaluate the sentence's Truth value depending on a following matching/mismatching picture. Consistently, the Event-Related Potentials (ERPs) and the RTs showed a significant interaction between the polarity and the truth value of the sentence (TPI): cost emerged in semantic integration by comparing affirmative true (AT) and affirmative false (AF) sentences whereas processing costs do emerged in the opposite direction for negative true (NT) compared to negative false (NF) sentences, as indicated by changes in N400 amplitudes (for a debate on N400 see Kutas & Federmeier, 2011; Saddy, Drenhaus, & Frisch, 2004; Van Berkum, Zwitserlood, Hagoort, & Brown, 2003). The authors concluded that processing of negative accordingly with the TSSH (Kaup et al., 2006b; Lüdtke, Friedrich, De Filippis, & Kaup, 2008a).

Nonetheless, Ludtke et al., (2008) design involved at least two confounds due to pragmatic felicity and target presence/absence, respectively. In fact, in the NT condition, they used sentence like *In front of the tower there is no lion* and presented the following picture with a ghost in front of a tower. Then, the behavioural costs in terms of Truth value verification and the N400 effects could be attributed to the fact that either no lion was present, or an unexpected ghost was present in the picture context. Scappini et al. (2015) implemented an SPV task as in the study described above but presented two scenes. Following a sentence like Asterix is not/is building a hut two pictures appeared. On one side a different character (e.g. Obelix) performed the named action (building a hut) and on the other side, Asterix did a different action (e.g. cutting wood) (Scappini, Vespignani, & Delfitto, 2015). In this design, both named subjects and actions were depicted in the overall scene, making the context pragmatically sounding for all comparisons (affirmative and negative, true and false conditions).

The authors replicated the Ludtke et al., (2008) TPI for RTs showing that AT and NF were faster to be evaluated than NT and AF. Scappini et al. (2015) overcame the problem of pragmatic felicity by always depicting the subject and the action named in the sentence. Nonetheless, further confounds might explain their results. Basically, the verification of the sentence-picture matching, in that case, could be performed by deploying attention selectively towards just one of the two side of the screen, since there were neither trials in which both characters did the same action nor trials in which the same character did two different actions. A picture in which Asterix was building a hut might be judged as mismatching by examining only the side of the screen in which Obelix was building a hut. This task-set could have suggested to the participants the strategy to systematically deploy attention toward a specific side of the screen (e.g. right) rather than checking what is depicted on both sides. If this strategy is employed AT sentences, fit better the strategy than AF where the target (the agent and the action cued by the sentence) is always present. In AT the scene shows both named character and action on the same side, while, in AF the named character and predicated action are on different sides. For the negative sentences, it is the opposite: for NT conditions the named character and action are on different sides of the screen whereas for NF conditions both the named character and action appear on the same side. The TPI obtained by Scappini et al. showing faster RT (and smaller N400) for AT and NF than AF and NT could thus be twofold. Indeed, the advantage could be due to: (a) the strategy to deploying attention on one of the two sides of the screen, and/or, (b) the two step-simulation hypotheses. Eventually, these interpretations are not incompatible. If part of the effect is due to a two-stages reasoning (b), this does not rule out that part of the advantage could also be due to differences in the deployment of visual attention (a).

The present study was set to disentangle these previous findings by investigating the early automatic stages of picture selection during sentence-guided search. The paradigm used by Ludtke et al., (2008) and Scappini et al., (2015) was adapted to focus on sentence verification and in particular on picture selection and in-depth picture identification after sentence cues, by means of early ERPs components such as N2pc (negative posterior contralateral wave after 180-300ms of the target onset) and SPCN (Sustained Posterior Contralateral Negative, 300600 ms). These two ERPs are respectively informative of the process of target selection and identification (Eimer, 1996; Mazza, Turatto, & Umiltà, 2007). Since N2PC and SPCN components are superimposed to visual exogenous components, they are susceptible to the complexity of the display. Then, the present study employed simple geometric shapes rather than cartoon characters as subjects of the sentences. For the same reason, it involved contradictory predicates referring to attribute of visual objects (i.e., shape and colour, Wolfe & Horowitz, 2004) rather than real-world actions. The SPV task allowed to more explicitly manipulate the predictability of the actual state of affair following a negation. It was expected to find a TPI showing a general advantage for AT and NF sentences compared to AF and NT sentences, respectively (Lüdtke et al., 2008b; Scappini et al., 2015). Across two experiments, it was manipulated also the number of alternatives in terms of representation of meaning in a context with multiple (three possible figures) and unique (only two possible figures) interpretation of the sentence. The aim was to extend previous results showing that contradictory predicates do not always require the representation of the negated state of affairs prior than the actual state of affairs (Spychalska et al., 2016; Tian & Breheny, 2016; Tian, Breheny, & Ferguson, 2010a). In an EEG study, Spychalska et al., (2016) challenged the TSSH by means of an SPV task by comparing a pragmatically ambiguous context offering multiple alternatives in negative meaning computation, to a pragmatically sounding context that pointed to a unique meaning. The authors found that in the unique alternative condition, the SPV of negative sentences was facilitated compared to the multiple alternatives condition, as indicated by N400 effects. Moreover, they found a similar advantage for affirmative and negative sentences in the unique alternative context. Interestingly, these results contrast with what predicted by the TSSH expecting that the comparison between multiple and unique cases should show higher cost for negative compared to affirmative sentences and they argue in favour of a more dynamic model of the representation of negative sentences. These results contrast with what predicted by the TSSH expecting that also such cases negative sentence verification should show higher cost compared to affirmative sentences (Tian & Breheny, 2016). Following such interpretation, in the present study, it was expected to find a reduced cost for negative sentences verification in a context in which the meaning of negative sentences could be easily computed (contradictory predicates) compared to a context with multiple alternatives (bipolar adjectives). It was expected not black to quickly point to white with respect to not rectangular to point a mental representation of circular when only two colours but more than two shapes exist in the context. Furthermore, the original contribution of this study relies on the N2pc component analysis that allow to focus on picture verification during sentence-guided search. The majority of ERPs studies on negation processing mainly investigated the modulation of the N400, an ERP component that has been shown to be a useful measure of semantic integration occurring later during sentence verification (Kutas & Federmeier, 2011). On the contrary, the N2pc component reflects the early stages of picture selection (visual target) and importantly, it allows for disambiguating which side of the screen the visual attention is directed to, in a spatial fashion (Eimer & Grubert, 2014).

4.4.1. Method: Experiment 1 SPV in a context with multiple alternatives

This study took advantage of search times and ERPs measures to investigate picture selection and sentence verification as a function of the sentence structure. Here, the sentence structure is defined mainly along two acceptations: polarity and adjective position. The former manipulation aimed to challenge the TPI by presenting contradictory and bipolar attributes in different positions. The investigation of the TPI and the adjective position offered an interesting insight into the question of how sentences and specifically, negative sentences trigger perceptual representation and the cascade effect on visual attention deployment.

Participants

Thirteen native Italian speakers (10 female; age = 24 years, SD = 3.9 years) took part in experiment 1. All of them had normal or corrected to normal vision and were reimbursed for their participation. All participants signed informed consent, and the experiment was conducted in accordance with the declaration of Helsinki and approved by the University of Trento's Ethics Committee. One participant was excluded due to less than 50% of trials kept due to artefacts in the EEG data. Thus, twelve participants were included in the analysis (10 females; mean = 24 years, SD = 3.6).

Stimuli, design, and Procedure

Experiment 1 comprised a Sentence Picture Verification (SPV) task that required a search task in which the target i.e. the shape and the colour mentioned in the sentence, were always present (see Figure 29). The participant was instructed to press a button (arrow up/down) to respond as to whether the sentence cue (see Figure 30) that predicated about shape and colour of specific figures matched. The target was indicated by a sentence that appeared word by word with an Intra Stimulus Interval (ISI) of 1500ms before the onset of the search array.

The cue was either an Affirmative or a Negative sentence (Polarity). Affirmative (1.a and 2.a) and Negative (1.b and 2.b) sentences were presented in random order in each block; there were 48 trials for each level of Polarity. Both cue, target and distractor objects were randomised between participants. All sentences were composed by a subject (i.e., Le Figure, The figures), a nominal adjective, for instance, Circular (i.e. 1, attributive position) and a non-verbal predicate - a copula and an adjective - for instance, black (i.e. 1, predicative position). In

Experiment 1, all sentences always showed a binary black/white and a ternary circular, triangular and rectangular adjective. Then, the experimental context had two possible colours and three possible shapes. The search array lasted for 200 ms always presenting the mentioned adjectives i.e. shape and colour.

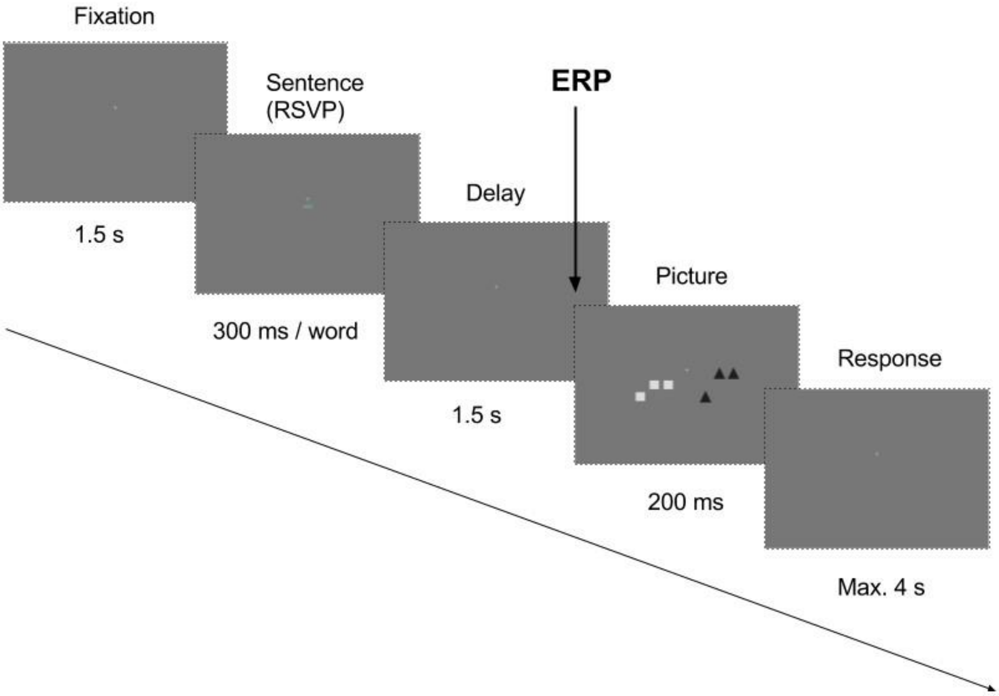


Figure 29. Paradigm of experiment one in which participants performed an SPV task. A fixation cross always preceded the cue by lasting 1500 milliseconds. Then, the sentence cue appeared word by word at the centre of the screen lasting 300 milliseconds each. Then, after a long delay of 1500 milliseconds, the picture array was presented for 200 milliseconds. The picture always showed two groups of figures with different shape and colour so as the shape and the colour mentioned in the sentence were always depicted. Following the offset of the picture, participant had a maximum of 4 seconds to indicate the Truth value of the sentence by pressing a button (arrow up vs arrow down).

(1) a. Le figure circolari sono nere
The figures circular$_{[+P]}$ are black$_{[+P]}$
The circular figures are black

b. Le figure circolari **non** sono nere
The figures circular$_{[+P]}$ **not** are black$_{[+P]}$
The circular figures are not black

(2) a. Le figure nere sono circolari
The figures black$_{[+P]}$ are round$_{[+P]}$
The black figures are circular

b. Le figure nere **non** sono circolari
The figures black$_{[+P]}$ **not** are circular$_{[+P]}$
The black figures are not circular

Figure 30. The linguistic stimuli were affirmative (1.a and 2.a) and negative (1.b and 2.b) sentences that always presented a nominal adjective for instance circular in 1 (i.e. attributive position) and a non-verbal predicate - a copula and an adjective - for instance circular in 2 (i.e. predicative position).

EEG recording and data analysis

The experiment was programmed in Matlab, using the Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997; Kleiner, Brainard, & Pelli, 2007). EEG was DC-recorded from 32 scalp electrodes at standard positions of the extended 10/20 system (frontal: Fp1, Fpz, F9, F7, F3, Fz, F4, F8, F10, FC5, FC6; central: T7, C3, Cz, C4, T8, CP5, CP6; parietal: P7, P3, Pz, P4, P8, PO7, PO8; occipital: O1, Oz, O2). One electrode was placed externally on each mastoid and below the left eye. All electrodes were referenced to the left mastoid and the ground was placed anteriorly to Fz. All electrodes impedances were less than 10 kΩ. The EEG was amplified and was filtered with a 40 Hz low-pass filter and then digitalised with a sampling rate of

1000 Hz and data. The continuous EEG was segmented from 100 ms to 600 ms relative to the onset of the search array. Trials with artefacts (horizontal EOG exceeding 30 µV, vertical EOG exceeding 60 µV, all other channels exceeding 80 µV) were removed. Waveforms for trials with correct responses were averaged for each condition separately.

N2pc and SPCN amplitudes were measured at electrodes PO7, and PO8 as ERP mean amplitudes between 180 and 300 ms and 300 to 600 ms post-stimulus, respectively. Jackknife-based analyses were used to compare onset latencies across tasks (Miller, Patterson, & Ulrich, 1998). N2pc onset is defined relative to an amplitude criterion of >=-1 µV. Finally, it was used the side of the display where the named shape was depicted as reference level. Negativity in the time windows of the N2pc (ca. 180-300 ms) and SPCN (ca. 300-600 ms) was, therefore, a measure of increased covert attention towards the named shape. This choice implies that the amplitude in the N2pc component reflects the amount of attention deployed towards the side of the screen in which the named shape is present. Positivity, on the contrary, would imply covert attention towards the named colour. For affirmative true and negative false conditions (e.g. a sentence cue as The black figures are circular or The black figures are not circular followed by black circles on one side and white rectangles on the other side) the side of named shape is also the side of named colour and it was expected a negativity in the N2pc irrespectively from the adjective position. In these cases, then, N2pc was not useful to disambiguate which feature was selected earlier. For affirmative false and negative true conditions (e.g. a sentence cue as The black figures are circular or The black figures are not circular followed by black triangles one side and white circles on the other side) it was expected the N2pc negativity to reflect early shape selection whereas the N2pc positivity to mirror colour selection because in both cases the named colour and shape were depicted on two opposite side of the screen. The collected EEG data were analysed using the software Brain Vision Analyzer (Brain Products GmbH, Germany) for the preprocessing and single subject averaging. The free software R was used for statistical analyses of both behavioural and ERP data.

Reaction times (RTs) of accurate responses as dependent variable were analysed by comparing Generalized linear

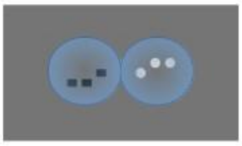
Mixed Effect Models, the model with the low dAIC and the higher AICw was selected as the best model.

4.4.2. Results

Behavioural results

All subjects answered in more than 88% of the trials correctly, and RTs of correct responses were included in the analysis, see Figure 31.

The model comparison among Generalized Linear Mixed Effects Models in Table 11, shows the multiplicative model to best approximate data (random effect are plotted in Appendix). The TPI interaction revealed that AT sentences predicted an advantage in terms of faster RTs compared to AF sentences ($\beta = 0.19$, SE = 0.018, t = 10.57). The contrary emerged for NT sentences that predicted slower RTs compared to NF sentences ($\beta = -0.11$, SE = 0.02, t = 6.27). Finally, the three-way interaction among Polarity, Truth value and Adjective position revealed that only in NT sentences verification, the shape in nominal position predicted faster RTs compare to the colour in nominal position ($\beta = 0.07$, SE = 0.01, t = 5.22) whereas Adjective position did not emerge to be a good predictor of RTs for other conditions.



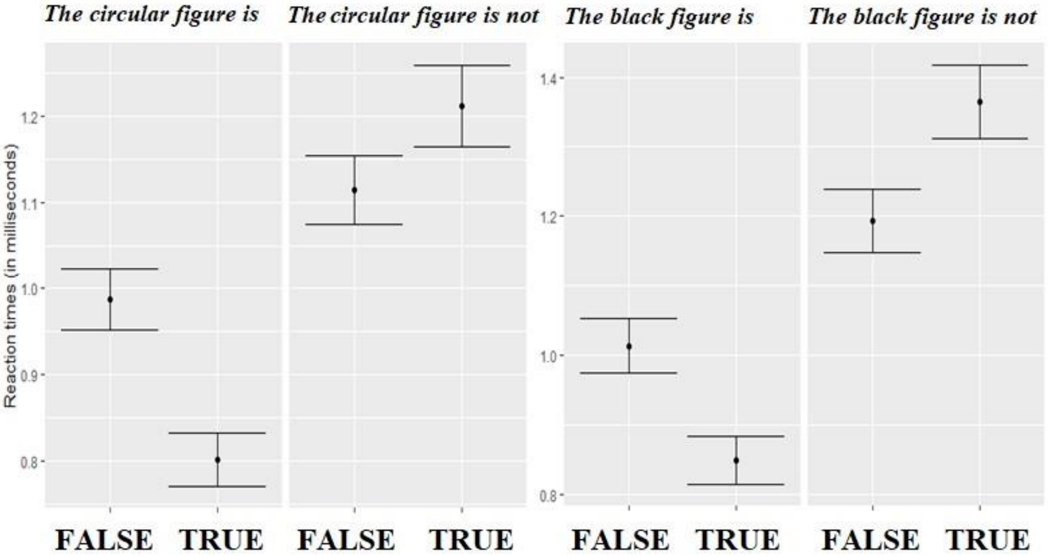


Figure 31. Mean reaction times (RTs) and 95% confidence intervals (CI) for ShapeColour on the right panels and Colour-Shape on the left panels split into Polarity, i.e. Affirmative and Negative cue sentences and further split into Truth value.

models	RD	dAIC	AICw	df	χ	p	η2
Reaction times [ms] ~ (1|Participants)	958	799	\	\	\	\	\
Reaction times [ms] ~ Polarity+ (1 | Participants)	855	212	0	1	588.34	<0.001	0.10
Reaction times [ms] ~ Truth value + (1 | Participants)	958	798	0	0	0	1	\
Reaction times [ms] ~ Adjective position + (1 | Participants)	952	768	0	0	30.70	<0.001	0.005
Reaction times [ms]~ Polarity * Truth value * Adjective position + (1 | Participants)	818	0	1	6	780	<0.001	0.13

Table 11. Model comparison for predicting RTs for sentence Polarity (Affirmative and Negative), Truth value (True and False) and for Adjective position (Shape in nominal position vs colour in nominal position). RD = residual deviance, dAIC = difference between a model's Aikake information criterion and those of the best model, AICw = AIC weight, df = degrees of

freedom of the chi-squared statistic, η2 = total eta squared as the ratio between the chi-squared and the residual deviance of the null model.

N2pc components

The Appendix shows ERPs at electrodes PO7/8 contralateral and ipsilateral to the target in all three cue conditions (Polarity, Truth value and Adjective position). Figure 32 shows the difference waveforms obtained by subtracting ipsilateral from contralateral ERPs. It is important to notice that the side of the screen where the named Shape was depicted, was used as reference, irrespectively of the condition. Thus, N2pc negativity reflects early attention deployment toward the side of the screen where the named shape was depicted whereas N2pc positivity indicates early attention selection towards the side of the screen depicting the named colour.

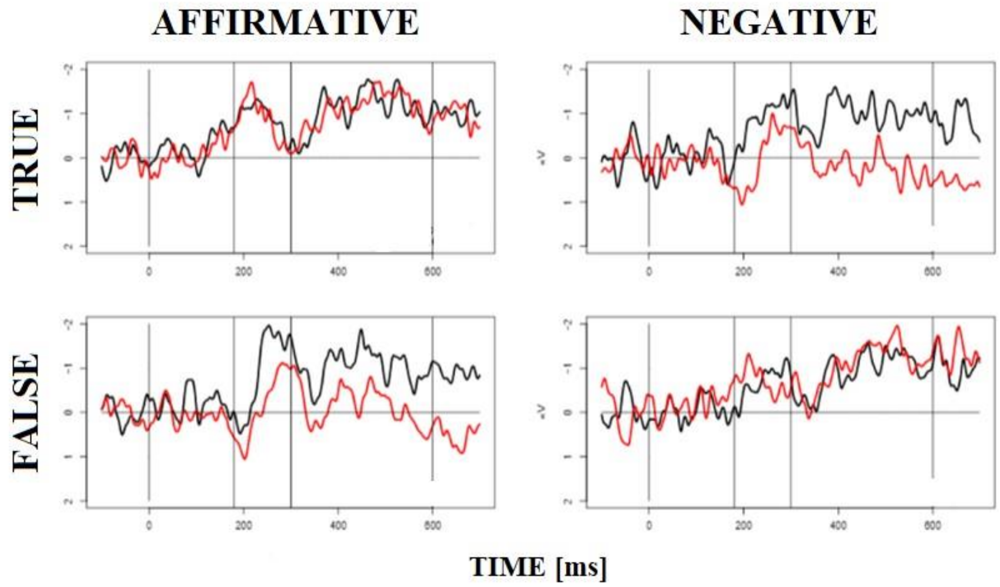


Figure 32. Difference waveform for Polarity and Truth value further split for Adjective position: Shape-Colour (black) and Colour-Shape (red). On the x-axis, the time course of average trial, the N2pc onset is expected about 180-200ms after the picture onset (0).

In table 12 is shown that model comparison that revealed the interactive model to be the best approximation to data. The results are similar to the RTs findings. In fact, the interaction revealed AT sentence to lead to higher amplitude in N2pc compared to AF sentences (β = 0.41, SE = 0.02, t = 20.16) and NF to lead to higher amplitude in N2pc compared to NT sentences (β = 0.58, SE = 0.02, t = 30.46). Finally, N2pc showed higher amplitude predicted by the adjective shape in nominal position compared to the adjective colour in nominal position, both for AF (β = 0.37, SE = 0.014, t =26.86) and NT (β = 0.37, SE = 0.01, t = 26.86) sentence verification. Whereas Adjective position emerged not to be a good predictor of RTs for AT and NF sentences. On note, it is possible to asses an early shape biased spatial attention deployment only in the AF and NT condition, where the named adjectives were depicted in different side of the visual array.

models	RD	dAIC	AICw	df	χ	p	η2
N2pc [mV] ~ (1|Participants)	859	274	\	\	\	\	\
N2pc [mV] ~ Polarity + (1 | Participants)	855	272	0	1	29.96	<0.001	0
N2pc [mV] ~ Truth value + (1 | Participants)	856	273	0	0	0	1	\
N2pc [mV] ~ Adjective position + (1 | Participants)	845	203	0	0	705	<0.001	0.0008
N2pc [mV]~ Polarity * Truth value * Adjective position + (1 | Participants)	837	0	1	6	2077	<0.001	0.002

Table 12. Model comparison for predicting N2pc amplitude in mV for sentence Polarity (Affirmative and Negative), Truth value (True and False) and for

Adjective position (Shape in nominal position vs colour in nominal position). RD = residual deviance, dAIC = difference between a model's Aikake information criterion and those of the best model, AICw = AIC weight, df = degrees of freedom of the chi-squared statistic, η2 = total eta squared as the ratio between the chi-squared and the residual deviance of the null model.

SPCN

Exactly as for N2pc analysis, the model comparison in Table 13 shows the interactive model to better approximate data of SPCN mean amplitudes in the 300-600 ms post-stimulus time window, See Figure 36. The findings were similar to the N2pc results. The interaction revealed AT sentence to lead to higher amplitude for the SPCN compared to AF (β = 0.48, SE = 0.03, t = -13.53) whereas the opposite emerged for negative sentences that prompted higher averaged SPCN amplitude for NF with respect to NT sentences (β = 0.60, SE = 0.03, t = 18.42). Finally, the adjective shape in nominal position lead to higher amplitude compared to the adjective colour in nominal position both for AF (β = 0.90, SE = 0.05, t = 18.32) and NT (β = 1.44, SE = 0.05, t = 28.75) sentences. Whereas it did not emerge to be a good predictor of RTs for AT and NF sentence verification.

models	RD	dAIC	AICw	df	χ	p	η2
SPCN [mV] ~ (1|Participants)	873	469	\	\	\	\	\
SPCN [mV] ~ Polarity + (1 | Participants)	872	444	0	1	32.77	<0.001	0.0001
SPCN [mV] ~ Truth value + (1 | Participants)	873	476	0	0	0	1	\
SPCN [mV] ~ Adjective position + (1 | Participants)	869	94	0	0	381.99	<0.001	0.001
SPCN [mV]~ Polarity * Truth value * Adjective position + (1 | Participants)	857	0	1	6	1132.632	<0.001	0.004

Table 13. Model comparison for predicting SPCN amplitude in mV for sentence Polarity (Affirmative and Negative), Truth value (True and False) and

for Adjective position (Shape in nominal position vs colour in nominal position). RD = residual deviance, dAIC = difference between a model's Aikake information criterion and those of the best model, AICw = AIC weight, df = degrees of freedom of the chi-squared statistic, η2 = total eta squared as the ratio between the chi-squared and the residual deviance of the null model.

4.4.3. Discussion Experiment 1

Participants read affirmative and negative sentences presented word by word at the centre of the screen while ERPs recording was locked at the visual display onset. Then, they were asked to verify whether a picture presented after a sentence matched or not in meaning. Under a non-incremental hypothesis of negation processing, RTs showed a TPI with False value being evaluated faster than True value only in Affirmative sentences while the opposite emerged in Negative sentences (Fischler, Bloom, Childers, Roucos, & Perry, 1983). The EEG results confirmed and extended the behavioural findings. According to the TSSH, both N2pc and SPCN amplitude revealed a significant TPI showing higher amplitude during AT compared to AF sentence verification and during NF compared to NT sentence verification. These results can be easily interpreted as an early efficient picture selection and in-depth identification, that led to a faster sentence verification during AT and NF sentences verification compared to their counterpart, respectively.

Nevertheless, these results may be better interpreted by considering the specific experimental context participants had to face. Basically, Experiment 1 offered multiple alternatives to compute negative sentence meaning. In fact, it involved three possible shapes, i.e. rectangular, circular and triangular. This manipulation was set to appreciate any difference in negation verification due to the demand of computing the target meaning in a context of multiple alternatives (Spychalska et al., 2016). That is, in Experiment 1 *not black* always meant white while *not rectangular* could either mean circular or rectangular. Then, the adjective position was a further manipulation that allowed to appreciate the interaction between the pragmatic ambiguity prompted by the context and the computation of the meaning of negative sentences. Actually, RTs showed a general advantage (short response times) when adjective shape was in nominal position e.g. *The rectangular figures are black*[8], compared to when shape was in predicative position e.g. *The black figures are rectangular*. This might reflect a strategy that participant likely employed in that context of multiple alternatives, rather to be a pure syntactic or semantic effect due to adjective position per se. That is, shape was a more diagnostic feature to get the target template (e.g., *not rectangular* had two competing alternatives). Thus, participants might prioritise the shape attribute to be salient in working memory to the detriment of the colour attribute. Hence, having the shape in nominal position and the colour in predicative position resulted easier with respect to the opposite condition. That is, in order to get a detailed target representation before verify the truth value of the sentences, the negative operator should not be applied to the shape adjective. This bias towards the information shape was clear-cut confirmed by looking at the N2pc polarity (i.e. negativity deflection attention toward the named shape vs positivity attention toward colour) which was preferentially deployed towards the display side where the named shape was depicted independently of all conditions. Finally, SPCN reflecting an in-depth target identification (Mazza et al., 2007) showed the same pattern of results, interpreted following the TSSH.

4.4.4. Experiment 2: SPV in a context with unique interpretation

The findings of Experiment 1 replicated the TPI as expected by the TSSH (Lüdtke et al., 2008a; Scappini et al., 2015). Moreover, the three-way interaction might induce concluding that the adjective position in negative sentences reduced the cost of the truth value verification depending on how many alternatives the context admitted to compute

---

[8] The sentences stimuli were in Italian that usually posit attributive adjectives in post-verbal position e.g. *The figures rectangular/black* instead of *The rectangular/black figures*, see Cinque, 2010.

(Spychalska et al., 2016). In order to test this possibility and to appreciate the degree by which the adjective position in negative sentences guide target selection, it was run a further experiment. in Experiment 2 participants were presented with a context in which the number of the target's features were balanced i.e. two shape and two colours. If Adjective position played a role in guiding picture selection during sentence verification then, it was expected to find a reduced cost in computing the meaning of negative sentences depending on which adjective was in nominal position (e.g. *The figure black…* or *The figure circular…*) with respect to the adjective in predicative position (e.g. *…is not circular* or *…is not black*), irrespectively from the conditions (for a further discussion about adjectives in nominal position see Cinque, 2010; Truswell, 2009). This manipulation was set also to disentangle whether the shape bias observed in Experiment 1 (i.e., N2pc negativity reflecting attention deployment toward the side depicting the named shape) was strictly dependent on the context demands or if it could be better placed in light of the role played by adjective position.

Participants

A new group of thirteen native Italian speaking participants (8 female; age: 26.31 years, SD: 5.88 years) took part in the experiment 2 all with normal or corrected to normal vision and were reimbursed for their participation. All participants signed informed consent, and the experiment was conducted in accordance with the Declaration of Helsinki and with the University of Trento's Ethics Committee. One participant was excluded due to less than 50% of trials kept due to artefacts in the EEG data. Thus, twelve participants were included in the analysis (7 females; mean = 25.92 years, SD = 5.95).

Stimuli, design, and Procedure

Methods of Experiment 2 were the same as in Experiment 1, apart from the unique interpretation offered by the context to compute the meaning of negative sentences, i.e., only two possible target's shapes (i.e. circular and rectangular) and colours (i.e. black and white).

4.4.5. Results

Behavioural results

All subjects answered in more than 90% of the trials correctly. Only the RTs relative to correct response of the truth value of the sentence were included in the analysis, see Figure 33. Table 14 shows the model comparison of GLMMs predicting RTs with the interactive model showing the best fit to RTs data. The TPI showed a different pattern of results compared to Experiment 1. AT did showed significant shorter RTs compared to AF (β = 0.22, SE =0.02, t = 12.26) sentences verification however, no difference emerged by comparing the NT to the NF sentences (β = 0.02, SE = 0.03, t = 0.60). The three-way interaction showed that shape in nominal position lead to faster RTs compared to colour in nominal position in AF (β = 0.07, SE = 0.03, t = 2.60), in NF (β = 0.11, SE = 0.04, t = 3.05) and NT (β = 0.13, SE = 0.04, t = 2.99) . However, Adjective position did not emerge as a good predictor of RTs for AT (β = 0.02, SE = 0.02, t =0.78) sentence verification.

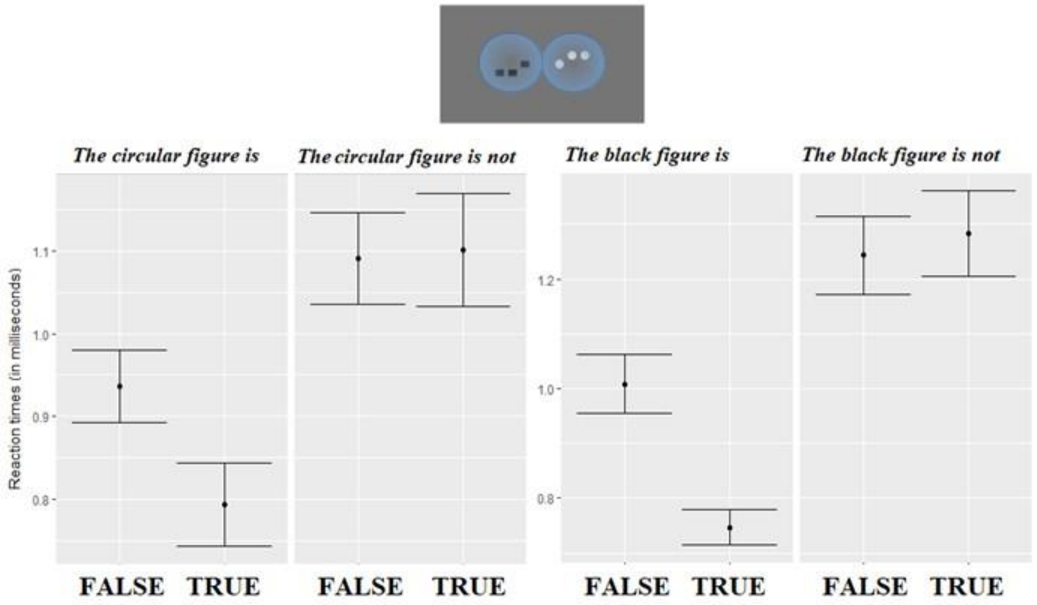


Figure 33. Mean reaction times (RTs) and 95% confidence intervals (CI) for ShapeColour on the right panels and Colour-Shape on the left panels split into Polarity, i.e. Affirmative and Negative cue sentences and further split into Truth value.

models	RD	dAIC	AICw	df	χ	p	η2
RTs [ms] ~ (1|Participants)	638	365	\	\	\	\	\
RTs [ms] ~ Polarity + (1 | Participants)	336	71	0	1	302.23	<0.001	0.06
RTs [ms] ~ Truth value + (1 | Participants)	604	339	0	0	0	1	\
RTs [ms] ~ Adjective position + (1 | Participants)	625	359	0	0	0	1	\
RTs [ms] ~ Polarity * Truth value * Adjective position + (1 | Participants)	222	0	1	6	402.32	<0.001	0.09

Table 14. Model comparison for predicting RTs for sentence Polarity (Affirmative and Negative), Truth value (True and False) and for Adjective position (Shape in nominal position vs colour in nominal position). RD = residual deviance, dAIC = difference between a model's Aikake information criterion and those of the best model, AICw = AIC weight, df = degrees of freedom of the chi-squared statistic, η2 = total eta squared as the ratio between the chi-squared and the residual deviance of the null model.

N2pc

The figure 34 shows the difference waveforms obtained by subtracting ipsilateral from contralateral ERPs in the N2pc time window 180-290 post-stimulus.

The model comparison in table 15 shows that the interactive model better approximated the averaged N2pc component amplitude. The TPI interaction showed that AT sentences prompted higher N2pc amplitude compared to AF sentence (β = 1.58, SE = 0.08, t = 20.45) the same pattern of results emerged for NT sentence that show a higher averaged N2pc amplitude compared to NF (β = 0.31, SE = 0.07, t = 4.17). Moreover, the three-way interaction showed that the adjective shape in nominal position lead to higher averaged N2pc in NT (β = 1.82, SE = 0.11, t = 16.18) and in NF (β = 0.39, SE = 0.09, t = 4.23) whereas Adjective position did not show to predict RTs in Affirmative sentences. On note, it is possible to asses an early shape-biased spatial attention deployment (negativity of N2pc) only in the AF and NT condition, where the named adjectives were depicted in different side of the visual array.

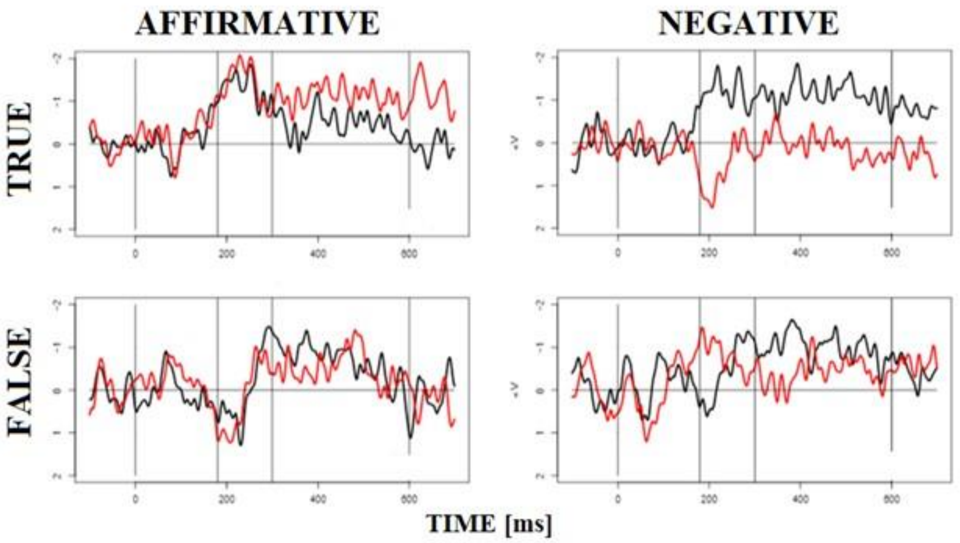
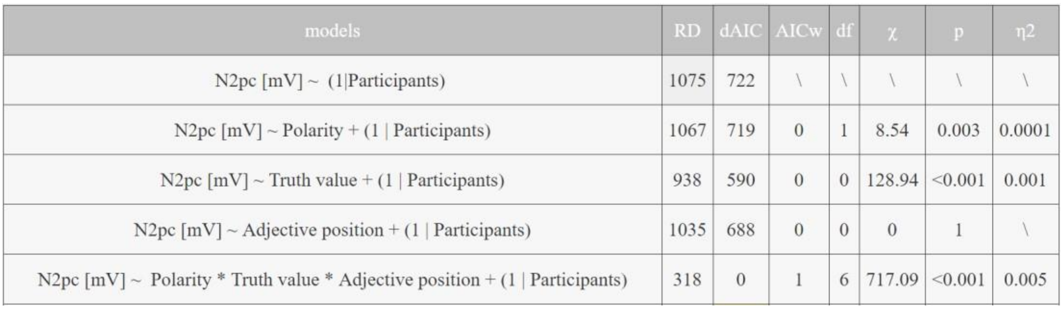


Figure 34. Difference waveform for Polarity and Truth value further split for Adjective position: Shape-Colour (black) and Colour-Shape (red). On the x-axis, the time course of average trial, the N2pc onset is expected about 180-200ms after the picture onset (0).



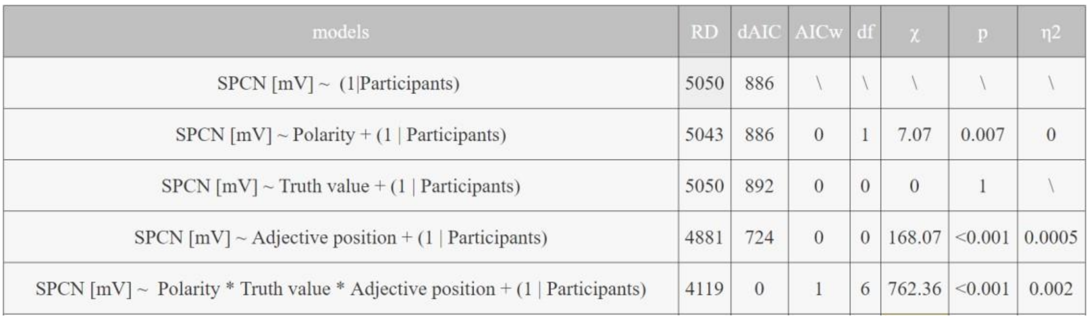
models	RD	dAIC	AICw	df	χ	p	η2
N2pc [mV] ~ (1|Participants)	1075	722	\	\	\	\	\
N2pc [mV] ~ Polarity + (1 | Participants)	1067	719	0	1	8.54	0.003	0.0001
N2pc [mV] ~ Truth value + (1 | Participants)	938	590	0	0	128.94	<0.001	0.001
N2pc [mV] ~ Adjective position + (1 | Participants)	1035	688	0	0	0	1	\
N2pc [mV] ~ Polarity * Truth value * Adjective position + (1 | Participants)	318	0	1	6	717.09	<0.001	0.005

Table 15. Model comparison for predicting N2pc amplitude in mV for sentence Polarity (Affirmative and Negative), Truth value (True and False) and for Adjective position (Shape in nominal position vs colour in nominal position). RD = residual deviance, dAIC = difference between a model's Aikake information criterion and those of the best model, AICw = AIC weight, df = degrees of freedom of the chi-squared statistic, η2 = total eta squared as the ratio between the chi-squared and the residual deviance of the null model.

SPCN

Exactly as for N2pc analysis, the model comparison in Table 16 shows the interactive model to better approximate data of SPCN mean amplitudes in the 300-600 ms post-stimulus time-window, See Figure 4. Results were similar to those on N2pc component. In fact, the TPI showed the AT leading to higher averaged SPCN amplitude compared to AF (β = 0.21, SE = 0.03, t = 5.96) the same pattern of results was found for NT sentences that predicted higher averaged SPCN compared to NF sentences (β = 0.17, SE = 0.04, t = 4.59). Finally, the three-way interaction revealed that the adjective shape in nominal position led to a higher averaged SPCN amplitude in all conditions namely, AT (β = 0.64, SE = 0.05, t

= 13.10), AF (β = 0.16, SE = 0.05, t = 3.27), NT (β = 1.24, SE = 0.05, t = 23.66) and NF (β = 0.59, SE = 0.05, t = 11.83).



models	RD	dAIC	AICw	df	χ	p	η2
SPCN [mV] ~ (1|Participants)	5050	886	\	\	\	\	\
SPCN [mV] ~ Polarity + (1 | Participants)	5043	886	0	1	7.07	0.007	0
SPCN [mV] ~ Truth value + (1 | Participants)	5050	892	0	0	0	1	\
SPCN [mV] ~ Adjective position + (1 | Participants)	4881	724	0	0	168.07	<0.001	0.0005
SPCN [mV] ~ Polarity * Truth value * Adjective position + (1 | Participants)	4119	0	1	6	762.36	<0.001	0.002

Table 16. Model comparison for predicting N2pc amplitude in mV for sentence Polarity (Affirmative and Negative), Truth value (True and False) and for Adjective position (Shape in nominal position vs colour in nominal position). RD = residual deviance, dAIC = difference between a model's Aikake information criterion and those of the best model, AICw = AIC weight, df = degrees of freedom of the chi-squared statistic, η2 = total eta squared as the ratio between the chi-squared and the residual deviance of the null model.

4.4.6. Discussion Experiment 2

In Experiment 2, participants were asked to verify the truth value of sentences in a pragmatically sounding context - in which negative sentences presented always contradictory predicates. The main aim of this manipulation was to disentangle the role played by the sentence structure - intended as sentence polarity and adjective position – during picture selection. The main result showed no cost due to any supposed representation of the negated state of affairs that the TSSH expects prior of the representation of the actual state of affairs, as the comparison of NT and NF sentences verification task (Tian & Breheny, 2016). AT lead to easier sentence-picture verification compared to AF sentences, whereas no differences resulted by comparing NT to NF sentences. These results might suggest that the representation of the negated state of affairs it is not a mandatory step during negation verification and that the representation of negative sentences changes depending on individuals knowledge about the context (Hasson & Glucksberg, 2006; Johnson-Laird, 1983; Spychalska et al., 2016; Tian et al., 2016). In a pragmatically sounding and no demanding context, the ERPs results showed that NT sentences leading to a triggered picture selection as shown by the N2pc component analysis. Furthermore, the N2pc polarity revealed that during Negative sentences verification visual attention was selectively deployed towards the side of the screen where the adjective in nominal position was depicted. Nonetheless, the adjective shape in nominal position facilitated sentence verification across conditions and preferentially biased attention towards the side where named shape was depicted by promoting a sustained, in-depth target identification compared to NF, as shown by the SPCN component analysis. These results suggest that the sentence structure intended as adjective position does guide visual attention deployment. Furthermore, they indicated that the visual features (shape and colour) cued by sentences affect the priority map of the template during guided-search similarly as pictorial cues does.

4.4.7. General conclusion

This study explored whether and how the sentence structure and the attribute of the target affects picture selection during sentence-guided search. Across two experiments, it replicated the cost shown by negative sentence during the SPV task. In the present study, the traditional result has been extended by the ERPs effects showing the ability of negative sentences to trigger efficient perceptual template. It seems that negating an attribute about an object does not provide any definite description of this object but still provide an efficient guide during picture selection, especially in pragmatically sounding contexts. The two clear-cut interactions between the demands of the context in terms of alternative representations and sentence structure in terms of adjective position posit the possibility to investigate intermediate logical, perceptual and linguistic representation during the SPV task. Future studies should (a) replicate these results by presenting sentence in rapid parallel visual presentation (RPVP, Snell & Grainger, 2017) according to the hypotheses and the evidence of long text processing, (b) disentangle the role of working memory capacity during negative sentence processing and (c) investigate hierarchy of the attribute of intermediate representations and, finally (d) such effects should be found during spoken sentence-picture verification task.

4.4.8. Bibliography

Carpenter, P. A., & Just, M. A. (n.d.). *Sentence comprehension: A psycholinguistic processing model of verification.*

Cinque, G. (2010). *The syntax of adjectives: A comparative study* (Vol. 57). MIT press.

Eimer, M. (1996). The N2pc component as an indicator of attentional selectivity. *Electroencephalography and Clinical Neurophysiology*, *99*(3), 225–234.

Eimer, M., & Grubert, A. (2014). The gradual emergence of spatially selective target processing in visual search: From feature-specific to object-based attentional control. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(5), 1819–1831. https://doi.org/10.1037/a0037387

Giora, R., Fein, O., & Aschkenazi, K. (n.d.). *Negation in Context : A Functional Approach to Suppression*. (March 2015), 37–41. https://doi.org/10.1080/01638530709336896

Giora, R., Heruti, V., Metuki, N., & Fein, O. (2009). "When we say no we mean no": Interpreting negation in vision and language. *Journal of Pragmatics*, *41*(11), 2222–2239.

Giora, R., Livnat, E., Fein, O., Barnea, A., Zeiman, R., & Berger, I. (2013). Negation generates nonliteral interpretations by default. *Metaphor and Symbol*, *28*(2), 89–115.

Hasson, U., & Glucksberg, S. (2006). Does understanding negation entail affirmation?: An examination of negated metaphors. *Journal of Pragmatics*, *38*(7), 1015–1032.

Haviland, S. E., & Clark, H. H. (1974). What's new? Acquiring new information as a process in comprehension. *Journal of Verbal Learning and Verbal Behavior*, *13*(5), 512–521.

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Harvard University Press.

Jordan, M. P. (1998). The power of negation in English: Text, context and relevance. *Journal of Pragmatics*, *29*(6), 705–752.

Just, M. A., & Carpenter, P. A. (1971). Comprehension of negation with quantification. *Journal of Verbal Learning and Verbal Behavior*, *10*(3), 244–253. https://doi.org/10.1016/S0022-5371(71)80051-8

Kaup, B., Lüdtke, J., & Zwaan, R. A. (2006a). Processing negated sentences with contradictory predicates: Is a door that is not open mentally closed? *Journal of Pragmatics*, *38*(7), 1033–1050. https://doi.org/10.1016/j.pragma.2005.09.012

Kaup, B., Lüdtke, J., & Zwaan, R. A. (2006b). Processing negated sentences with contradictory predicates: Is a door that is not open mentally closed? *Journal of Pragmatics*, *38*(7), 1033–1050.

Kaup, B., & Zwaan, R. A. (2003). Effects of negation and situational presence on the accessibility of text information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(3), 439–446. https://doi.org/10.1037/0278-7393.29.3.439

Khemlani, S., Orenes, I., & Johnson-Laird, P. N. (2012). Negation: A theory of its meaning, representation, and use. *Journal of Cognitive Psychology*, *24*(5), 541–559.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*, 621–647. https://doi.org/10.1146/annurev.psych.093008.131123

Luck, S. J., & Hillyard, S. A. (n.d.). Spatial filtering during visual search: Evidence from human electrophysiology. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(5), 1000–1014. https://doi.org/10.1037/0096-1523.20.5.1000

Lüdtke, J., Friedrich, C. K., De Filippis, M., & Kaup, B. (2008a). Event-related potential correlates of negation in a sentence-picture verification paradigm. *Journal of Cognitive Neuroscience*, *20*(8), 1355–1370. https://doi.org/10.1162/jocn.2008.20093

Lüdtke, J., Friedrich, C. K., De Filippis, M., & Kaup, B. (2008b). Event-related potential correlates of negation in a sentence-picture verification paradigm. *Journal of Cognitive Neuroscience*, *20*(8), 1355–1370. https://doi.org/10.1162/jocn.2008.20093

Mayo, R., Schul, Y., & Burnstein, E. (2004). "I am not guilty" vs "I am innocent": Successful negation may depend on the schema used for its encoding. Journal of Experimental Social Psychology, 40(4), 433-449.

Mazza, V., Turatto, M., & Umiltà, C. (2007). *Attentional selection and identi W cation of visual objects are re X ected by distinct electrophysiological responses*. 531–536. https://doi.org/10.1007/s00221-007-1002-4

Mazza, V., Turatto, M., Umiltà, C., & Eimer, M. (2007). Attentional selection and identification of visual objects are reflected by distinct electrophysiological responses. *Experimental Brain Research*, *181*(3), 531–536. https://doi.org/10.1007/s00221-007-1002-4

Nako, R., Wu, R., Smith, T. J., & Eimer, M. (2014). *Item and Category-Based Attentional Control During Search for Real-World Objects : Can You Find the Pants Among the Pans ? 40*(4), 1283–1288.

Ort, E., Fahrenfort, J. J., Ten Cate, T., Eimer, M., & Olivers, C. N. L. (2019). Humans can efficiently look for, but not select multiple visual objects. *BioRxiv*, 653030.

Pashler, H. (1987). Target-distractor discriminability in visual search. *Perception & Psychophysics*, *41*(4), 285–292.

Saddy, D., Drenhaus, H., & Frisch, S. (2004). Processing polarity items: Contrastive licensing costs. *Brain and Language*, *90*(1–3), 495–502.

Scappini, M., Vespignani, F., & Delfitto, D. (2015). *On the non-incremental processing of negation: A pragmatically licensed sentence-verification study with Italian adults*.

Snell, J., & Grainger, J. (2017). The sentence superiority effect revisited. *Cognition*, *168*, 217–221.

Spychalska, M., Haase, V., Kontinen, J., & Werning, M. (n.d.). *Processing of affirmation and negation in contexts with unique and multiple alternatives: Evidence from event-related potentials*.

Team, R. C. (2018). *R: A language and environment for statistical computing; 2015*.

Tettamanti, M., Manenti, R., Della Rosa, P. A., Falini, A., Perani, D., Cappa, S. F., & Moro, A. (2008). Negation in the brain: Modulating action representations. *Neuroimage*, *43*(2), 358–367.

Tian, Y., & Breheny, R. (n.d.). *Dynamic Pragmatic View of Negation Processing*. https://doi.org/10.1007/978-3-319-17464-8

Tian, Y., Breheny, R., & Ferguson, H. J. (2010a). *Rapid communication Why we simulate negated information : A dynamic pragmatic account*. *63*(12), 2305–2312. https://doi.org/10.1080/17470218.2010.525712

Tian, Y., Breheny, R., & Ferguson, H. J. (2010b). Why we simulate negated information: A dynamic pragmatic account. *The Quarterly Journal of Experimental Psychology*.

Tian, Y., Ferguson, H., & Breheny, R. (2016). *Processing negation without context why and when we represent the positive argument*. *3798*. ttps://doi.org/10.1080/23273798.2016.1140214

Truswell, R. (2009). Attributive adjectives and nominal templates. *Linguistic Inquiry*, *40*(3), 525–533.

Van Berkum, J. J. A., Zwitserlood, P., Hagoort, P., & Brown, C. M. (2003). When and how do listeners relate a sentence to the wider discourse? Evidence from the N400 effect. *Cognitive Brain Research*, *17*(3), 701–718.

Wolfe, J. M., & Horowitz, T. S. (2004). *What Attributes Guide the Deployment of Visual Attention and How Do They Do It?* (July). https://doi.org/10.1038/nrn1411

5. CONCLUSION

The present project aimed to investigate the degree at which language-mediated representations influence visuospatial attention deployment, in infants and adults. The theoretical framework embraced a developmental perspective that assumes the adult cognitive system to be the (temporary) ending point of a nonlinear pathway traced by the early developing brain (Chapter 1). Concurrently, it encompassed the possibility to investigate the language-cognitive system through candidate domain-relevant mechanisms such as disengagement of attention and target selection. These mechanisms are good candidates to study the memory-attention interface because they operate across domains of knowledge (visual-space and auditory-time). Moreover, the same kind of mental representation (multisensorial, linguistic-mediated) differently affects those visuospatial mechanisms as a function of time and experience. Perceptual and linguistic representations have been traditionally studied in separate rooms and employing models shaped on the adult cognitive system. However, a growing number of studies is trying to reconcile the theoretical and methodological debates typical of each different approach by adopting a developmental perspective (Chapter 1, 2 and 3). Nonetheless, the subject of investigation of different cognitive scientists is often the same act of cognition but studied through different lenses.

5.1. Theoretical issues

In sections 4.1 and 4.2, through indirect measures of resource allocation and visual engagement, two studies investigated the visual encoding of novel objects that could be paired with linguistic and nonlinguistic auditory sounds. This manipulation had the further goal to investigate the cascade effects of subsequent disengagement of attention mechanism as a function of the flexible retrieval of the representation of the novel object previously encoded. From a methodological point of view, the Training-Test paradigm allowed to use novel stimuli that did reduce the familiarity confound necessarily introduced in study that involves real stimuli and the GLMMs analysis - that account for individual variability by estimating participants random effects - did helped to reduce the error carried out by early individual bias towards specific object attributes. From a theoretical point of view, the comparison between the patterns of attentional deployment during each familiarisation with novel stimuli revealed an opposite trend between 12month-olds and adults. Indeed, infants revealed multisensory representation to trigger better encoding of new stimuli compared to visual information only, whereas the opposite pattern emerged for adults (for similar results in children and adults, see Matusz, Merkley, Faure, & Scerif, 2019). However, the resource allocation during disengagement of attention in the following visuospatial task revealed that both infants and adults made similar use of multisensory representations during subsequent attention deployment. Indeed, the adults' study replicated the facilitatory label effect expected by the Label Feedback Hypothesis: labels triggered disengagement of attention compared to both auditory and visual representations only. This effect appeared to be more strongly associated with an on-line facilitatory effect of the label, in adults; whereas infants' descriptives data showed a general advantage lead by multisensorial representation compared to visual information only, independently from on-line presentation of labels, by suggesting that language-mediated representation encoded in memory might lead to more stable and flexible representation guiding subsequent behaviour, during infancy. The results found here do contrast those of Twomey and Westermann (2018) who found that familiar audio-visual stimuli presented silently triggered a 'novelty response', as shown by longer looking times compared to familiar visual stimuli. In contrast, here it was found a 'consistency response' during the presentation of a familiar audio-visual stimulus. Indeed, when presented silently, it prompted resources allocation required for disengagement of attention, compared to the presentation of a visual object which was familiarised silently.

The results of these two studies replicated converging evidence of a preferential sensitivity to multisensorial stimuli in infants compared to adults (Chapter 2). Once again, this replication stresses the importance of mental ages comparison as a preferential gate to take charge of the understanding of the whole language-cognitive system. The comparison among mental ages allows getting a theory about language and visuospatial functions that account for time-dependent and experience-dependent influences and to detect those diagnostic markers of atypical visuospatial and language development.

Indeed, on-line and memory-based top-down mechanisms did guide the performance across the two groups by showing that the language system plays a pervasive role in sensory and attentional processes, since infancy. Furthermore, the consistency effect found across the two studies suggests that different mechanisms are dedicated to the processing of the object colour and label; hence, the deprivation of such features differently affects attention control. Furthermore, colours and labels are objects' attribute with cascade pervasive consequences for visuospatial tasks such as overlap and visual search.

Section 4.3 reported two visual searches challenging the specific role of words as categorical cues using visuospatial ERPs components sensitive to template guided search mechanisms. The categorisation ability was defined by the distinction between perceptual and conceptual distinctiveness. The object's features of an item referred to levels of category classification. These levels that entails the activation of perceptual features are likely to be category-specific (e.g., fruits vs stationery) and to be acquired through visual experience. Thanks to language, the categorisation ability overcomes the boundary between perceptual and conceptual features and represents a good case to investigate the relationship between linguistic-mediated representations and visuospatial attention mechanisms. The comparison between word and pictorial cues during two category-guided searches showed that labels triggered perceptual template preferentially biased towards high conceptual and perceptual distinctive categories, as perceptual cues do. These results suggest that in such ecological search tasks, among items of different categories, perceptual similarity does not fully explain behaviour variability.

On the contrary, linguistic mediated representation is likely a needed step during category-based search to compute the semantic framework that allows to perform the task and to trigger canonical conceptual and perceptual template. Indeed, in our every-day life, labels are frequently used as cues to visual objects instead of pictorial information to facilitate target selection. Thinking of the cognitive system as a necessarily linguistic system should, therefore, challenge future study on visual attention mechanisms.

Finally, Chapter 4.4. took advantage of visuospatial ERPs to investigate the early deployment of attention toward target features triggered by the sentence structure, during the computation of the sentence pragmatic value. Structured information conveyed by a sentence unit allowed to re-prioritize the features map of the pictorial template that guides the sentence-picture verification by overcoming the line of domain-specific mechanisms. The results showed that the strategies to compute the meaning cued by sentences strictly depended on the alternatives offered by a specific context (multiple or unique alternatives). Indeed, in the context of multiple alternatives, a two-step hypothesis of negative sentence representation fits the interpretations of Truth per Polarity Interaction found both in behavioural and ERPs data.

Nevertheless, by reducing the degree of freedom of such alternatives negative sentence seems to be represented in the absence of any simulation of the negated state of affair, as shown by the behavioural results that did not show any disadvantage for negating true sentence compared to negative false sentences.

Furthermore, the pragmatic effect of the context array interacted with the adjective position in sentences cue. This last manipulation was set to explicitly manipulate the features priority maps triggered by adjectives in nominal position compare to those presented postverbally. In fact, in the context of multiple alternatives, the perceptual saliency of shape did guide early attention deployment systematically toward the side of the screen where the mentioned shape was depicted (multiple alternatives for shape computation), independently from the adjective position. However, by reducing ambiguity in the context of unique alternatives, the early target selection emerged to be guided by those adjectives presented in nominal position. Nevertheless, further analyses are required to asses the statistical difference between the two experiments.

Finally, the feature shape showed a facilitatory effect in guiding a faster search task overall experiments. These results interestingly show that sentence trigger stable and flexible perceptual representation which saliency map depends on the priority of attributes of both linguistic (adjective position) and pictorial (shape) domains. Thus, sentence structure does matter during the template-guided search, and it implicitly directs the attention focus towards those information unit that are prioritised in the dimension map triggered by language-mediated representations.

5.2. Methodological issues

This thesis tried to bridge different approaches (visuospatial attention, word comprehension and sentence verification) and various techniques (behavioural, eye-tracking and EEG measures) to stress the investigation of the cognitive system representation ability in tasks able to disentangle effects domain-dependent (visual search) and domain-independent (orienting of attention) and to coherently interpreted multiple source of information that are clues to understand human behaviour and cognitive functioning. A multi-method approach was necessary in order to investigate language-cognition interplay independently from the experimental context and more importantly, across infant and adult population. Converging evidence coming from multiple sources of investigation do contribute to the possibility to generalize findings independently from specific controlled contexts. Moreover, converging evidence allows ascribing reliable, functional interpretation of the linguistic-cognitive system employing a wide range of candidate measures of specific linguistic and visuospatial mechanisms. Although it is a hard task to reconcile the scientific debates characterising the investigation of each separate mechanism, an integrated approach is worth it, for at least three reasons: (1) question-oriented multimethod approaches promote a fruitful communication and collaboration among researchers with different skills and background by leading to the computation of new models of cognition that can take charge of the whole linguistic cognition system, by addressing findings coming from different sub-fields; (2) to acquire knowledge about different methodologies means to be able select those methods that better allow to answer specific theoretical question rather than been constrained by conservative attitude in persevering on the very same line of research that reduce reliability and generalizability of findings, and that proceeds at its own pace neglecting more promising ways that suit the question under discussion in favour of the easiest way to run (and publish) a study. Especially in such a multivariate phenomenon like cognition. In conclusion, (3) a multi-method approach allow to get new insights on the traditional explanation of a specific phenomenon of interest in a specific discipline and also, it allows dressing the lens of other researchers that investigate the same act of cognition (and sometimes find similar outcomes) from different perspective and through different theoretical explanations.

5.3. Strengths and further perspectives

During these last three years of the doctoral school, I have been necessary in touch with experts from different approaches: developmental, language and visual attention scientists. Without such confrontations, the present work would not have been possible. It has been suddenly crystal clear from me that it exists a gap between the sincere interest of the heterogeneous cognitive science community to connect the dots traced by each branch of cognitive science in a coherent scenario able to specify an informative model of human cognition, and the conservative attitude to persevere in investigating the very same mechanism in a static approach that by-pass the whole picture in favour of reductionists explanation of separate mechanisms that do not always replicate across experimental contexts both within the same approach and among approaches. This project was motivated by the need for a shared common ground in cognitive science that should include psychologists, linguists, computer scientists and philosophers all interested in understanding human behaviour, in a time-dependent fashion. Here it is claimed that further perspective on language and visuospatial investigation should account for the interaction of these functions also investigating performance in tasks that have been traditionally implemented to investigate separated mechanisms. Further studies should firstly, replicate the evidence found in this project that shows a pervasive effect of language-mediated representation on visuospatial attention and should stress the theoretical questions about the role of meaningful labels further structured in sentences (noun-adjective, cleft sentences, polarity etc) in performing tasks relegated to memory and attention studies. The general expectation for a new line of integrated research should expect language to exert its generative power at the very early stage of cognitive development.

LIST OF FIGURES

Figure 2. Picture from Logan (2002), A family tree expressing the relations between the ancestors of the attention branch and those of the memory branch. Theories connected by bidirectional arrows can be made equivalent to each other, through ITAM. FIRM = fixed-capacity independent race model; TVA = theory of visual attention; CODE = COntour Detector theory of visual attention; ECTVA = executive control of TVA; GCM = generalized context model; EBRW = exemplar-based random walk model; EGCM = extended generalized context model; ALCOVE = attention learning COVEring theory; RASHNL = rapid attention Shifts 'N' Learning theory; ITAM = instance theory of attention and memory........7

Figure 3. The LFH model (Lupyan, 2012). A schematic of a neural network architecture for exploring on-line effects of labels on perceptual representations........8

Figure 5. Interaction of language and experience (Gentner & Boroditsky, 2001)........26

Figure 6. The Division of Dominance (Gentner, 1982)........28

Figure 7. Procedure by Sloutsky and Robinson (2008). 30

Figure 8. The Posnerian attention from Raz (2004). A sketch of the functional anatomy of the attentional networks. The pulvinar, superior colliculus, superior parietal lobe, and frontal eye fields are often activated in studies of the orienting network. The temporoparietal junction is active when a target occurs at a novel location........42

Figure 9. A gap-overlap task where a target can occur after fixation offset (gap), with the fixation remaining on-screen (overlap) from Keehen et al., 2013........47

Figure 10. The autonomic system's parasympathetic and sympathetic pathways involved in pupillary control from Patwari et al., 2012. 49

Figure 11. Participants were exposed to a) two blocks, namely the Word-Object (WO) and the Object Only (OO) block in a counterbalanced order. Both blocks started with (b) a training phase in which an object was present paired with a word or silently, respectively. Right after the training phase, (c) an overlap task started, and after a fixation cross lasting for 500 ms, the previously trained object was presented for 1000ms. Each overlap task comprised two conditions, namely a Consistent (where the object was identical to the one previously trained) and an Inconsistent in which the object was presented without a feature either the word in Word-Object or the colour in Object Only...73

Figure 12. Total looking times in milliseconds during the Object Only and the Word-Object training; the plot reports the means and the 95% Confidence Interval (CI) for the two trainings. 78

Figure 13. Baseline corrected pupil size in millimetres across time during the Training (on the left) and the following Overlap task's conditions Consistent and Inconsistent (on the right). The mean effect and the 95% Confidence Intervals (CI) are plotted..79

Figure 14. Interaction plot of Saccadic latency during the Overlap task's conditions further split for Object Only in red and Word-Object in green. Mean values and 95% CI are plotted...81

Figure 15. Effect plot of GLMM models predicting looking times in ms during the Word-Object training (on the left panel) and saccadic latency during the Consistent condition in the Word-Object block (on the right pane) for the Lexical Comprehension quotient (Lexical age/Chronological age) of the PVB questionnaire...82

Figure 16. Participants were asked to pay attention to two a) training phases consisting of 9 trials (2 seconds each) in two different blocks (Word-Object

and Object Only). Following each training participants performed the b) ORt, if they reached at least 80% of accuracy then they performed c) the Overlap task in which after a fixation cross (500 ms) a central object (S1) was presented in three conditions (Consistent, Inconsistent and Deprived) for 1 second, then a peripheral stimulus (S2) could randomly appear on the right or on the left of S1, at the same degree of eccentricity. 99

Figure 17. Baseline-corrected pupil size in millimetres across Time in milliseconds during both the Object Only (red) and the Word-Object (green) Training in the left panel; and during the Overlap task' conditions (Consistent, Inconsistent and Deprived) after the Object Only and the Word-Object training, on the right panel. Plotted the max and minimum values, the mean effect (bold line) and the 95% Confidence Intervals (CI, dark grey). 104

Figure 18. Interaction plot of Saccadic latency during the Overlap task's conditions further split for Object Only and Word-Object. Mean values and SE are plotted. 106

Figure 19. Box plot and dot plot of the computed On-line word effect (Consistent Object Only – Consistent Word-Object) and Afterword effect (Consistent Object Only – Deprived Word-Object) in milliseconds, for each subject. First, second and third quartile are plotted. The black line indicates the facilitation effect of word (>0) 108

Figure 20. Baseline-corrected pupil size in millimetres across Time in milliseconds during each trial of both the Tone (red) and the Word (green) training and across the Overlap task's conditions (Consistent and Inconsistent) in the Tone-Object and the Word-Object block. Plotted the mean effect (bold line) and the 95% Confidence Intervals (CI, dark grey). 112

Figure 21. Interaction plot of Saccadic latency during the Overlap task's conditions further split for Tone-Object and Word-Object blocks. Mean values and SE are plotted 114

Figure 22: Paradigm of experiment one in which participants performed a visual search task. A fixation cross always preceded the cue by lasting 800 milliseconds. Then, the cue appeared and lasted for 300milliseconds. The three possible cues (exact-match, picture and word) were presented in a blocked design (ABAB) Following the interval intra stimulus (ISI) of 200 milliseconds, the search array always showed the target among three distractors. The search array stayed visible until the participant indicated by pressing a button (arrow up vs arrow down) where the target was presented. 127

Figure 23. The first plot shows mean correct reaction times (RTs) and the SE and the second plot shows error rate for all trials conditions. Means and error bars are plotted........ 130

Figure 24. In the upper side: ERPs at electrodes PO7/8 contralateral and ipsilateral to the target in all three cue conditions Picture, Exact-match and Word, and difference waveforms obtained by subtracting ipsilateral from contralateral ERPs. In the downside: Difference waveforms for all the cue conditions Picture and Word separated into Canonical and Non-canonical. ... 134

Figure 25. Paradigm of experiment 2 in which participants performed a categorical search task. A fixation cross always preceded the cue by lasting 800 milliseconds. Then, the cue appeared and lasted for 200milliseconds. The three possible cues a) supraordinate category, b) basic-level category and c) picture were presented in a blocked design (ACBACB). Following an (ISI) of 300 milliseconds, the search array always showed the target among three distractor objects. The search array stayed visible until participant indicated the target position by pressing a button (arrow up vs arrow down)................. 139

Figure 26. Interaction plots of correct RTs (on the left panel) and accuracy rate (correct trials/ total trials) on the right panel for Cue Conditions (picture, basic level and supraordinate level) further split for Canonical (red) and Non-canonical (green) shape of category-items. Standard errors are plotted........... 139

Figure 27. Difference wave Difference waveforms of grand averaged N2pc amplitude for all the Cue Conditions (Picture, Basic-level and Supraordinate) separated into Canonical and Non-canonical. 141

Figure 28. Interaction plot of the averaged N2pc amplitude 190/290 milliseconds post-stimulus for the Canonical and Non-canonical shape of category-items further split for Cue Conditions (supra-ordinate category, basic-level category and picture). Standard errors for repeated measures are plotted. 142

Figure 30. The linguistic stimuli were affirmative (1.a and 2.a) and negative (1.b and 2.b) sentences that always presented a nominal adjective for instance circular in 1 (i.e. attributive position) and a non-verbal predicate - a copula and an adjective - for instance circular in 2 (i.e. predicative position). 161

Figure 31. Mean reaction times (RTs) and 95% confidence intervals (CI) for ShapeColour on the right panels and Colour-Shape on the left panels split into Polarity, i.e. Affirmative and Negative cue sentences and further split into Truth value. 164

Figure 32. Difference waveform for Polarity and Truth value further split for Adjective position: Shape-Colour (black) and Colour-Shape (red). On the x-axis, the time course of average trial, the N2pc onset is expected about 180-200ms after the picture onset (0). 166

Figure 33. Mean reaction times (RTs) and 95% confidence intervals (CI) for ShapeColour on the right panels and Colour-Shape on the left panels split into Polarity, i.e. Affirmative and Negative cue sentences and further split into Truth value. 172

Figure 34. Difference waveform for Polarity and Truth value further split for Adjective position: Shape-Colour (black) and Colour-Shape (red). On the x-axis, the time course of average trial, the N2pc onset is expected about 180-200ms after the picture onset (0). 174

LIST OF TABLES

Table 1. Model comparison for predicting Looking times in the Training phase. RD = residual deviance, AIC = Aikake information criterion, dAIC = difference between a model's AIC and those of the best model, AICw = AIC weight, df = degrees of freedom of the chi-squared statistic, η2 = eta squared as the ratio between the chi-squared and the residual deviance of the null model……………………………………………………………………………………………78

Table 2. Model comparison for predicting Saccadic latency in the Overlap tasks, RD = residual deviance, AIC = Aikake information criterion, dAIC = difference between a model's AIC and those of the best model, AICw = AIC weight, df = degrees of freedom of the chi-squared statistic. η2 = eta squared as the ratio between the chi-squared and the residual deviance of the null model……………………………………………………………………………………………79

Table 3. Model comparison for predicting Pupil dilation in the Training phase. RD = residual deviance, AIC = Aikake information criterion, dAIC = difference between a model's AIC and those of the best model, AICw = AIC weight, df = degrees of freedom of the chi-squared statistic, η2 = eta squared as the ratio between the chi-squared and the residual deviance of the null model………………..………………………..……………………………..………………………..104

Table 4. Model comparison for predicting Saccadic latency during the Overlap tasks' condition after both the Object Only and the Word Object training. RD = residual deviance, AIC = Aikake information criterion, dAIC = difference between a model's AIC and those of the best model, AICw = AIC weight, df = degrees of freedom of the chi-squared statistic, η2 = eta squared as the ratio between the chi-squared and the residual deviance of the null model……………………………………………………………………………………………106

Table 5. Model comparison for predicting difference (ð) in Saccadic latency by the on-line word effect (Consistent OO – Consistent WO) and the after-word effect (Consistent OO – Deprived WO). RD = residual deviance, AIC = Aikake information criterion, dAIC = difference between a model's AIC and those of the best model, AICw = AIC weight, df = degrees of freedom of the chi-squared statistic, η2 = eta squared as the ratio between the chi-squared and the residual deviance of the null model……………………………………………………………………………………108

Table 6. Model comparison for predicting Pupil dilation in millimetre during the Overlap task's conditions in the TO and the WO blocks. RD = residual deviance, AIC = Aikake information criterion, dAIC = difference between a model's AIC and those of the best model, AICw = AIC weight, df = degrees of freedom of the chi-squared statistic, η2 = eta squared as the ratio between the chi-squared and the residual deviance of the null model……………………………………………………………………………………112

Table 7. Model comparison for predicting Saccadic latency in milliseconds during the Overlap task's conditions in the TO and in the WO blocks. RD = residual deviance, AIC = Aikake information criterion, dAIC = difference between a model's AIC and those of the best model, AICw = AIC weight, df = degrees of freedom of the chi-squared statistic, η2 = eta squared as the ratio between the chi-squared and the residual deviance of the null model……………………………………………………………………………………113

Table 8. Model comparison for predicting RTs for Cue Condition (Exact-match, Picture and Word) and Shape (Canonical vs Non-canonical). RD = residual deviance, dAIC = difference between a model's Aikake information criterion and those of the best model, AICw = AIC weight, df = degrees of freedom of the chi-squared statistic, η2 = total eta squared as the ratio between the chi-squared and the residual deviance of the null model……………………………………………………………………………………130

Table 9. Model comparison for predicting Accuracy as the proportion of correct answer on the total trials, for the Cue Condition (Exact-match, Picture and Word) and Shape (Canonical vs Non-canonical). RD = residual deviance, dAIC = difference between a model's Aikake information criterion and those of the best model, AICw = AIC weight, df = degrees of freedom of the chi-squared statistic, η2 = total eta squared as the ratio between the chi-squared and the residual deviance of the null model……………………………………………………………………………131

Table 10. Model comparison for predicting Amplitude N2pc for the Cue Condition (Exact-match, Picture and Word) and Shape (Canonical vs Non-canonical). RD = residual deviance, dAIC = difference between a model's Aikake information criterion and those of the best model, AICw = AIC weight, df = degrees of freedom of the chi-squared statistic, η2 = total eta squared as the ratio between the chi-squared of each model and the residual deviance of the null model……………………………………………………………………………133

Table 11. Model comparison for predicting RTs for sentence Polarity (Affirmative and Negative), Truth value (True and False) and for Adjective position (Shape in nominal position vs colour in nominal position). RD = residual deviance, dAIC = difference between a model's Aikake information criterion and those of the best model, AICw = AIC weight, df = degrees of freedom of the chi-squared statistic, η2 = total eta squared as the ratio between the chi-squared and the residual deviance of the null model……………………………………………………………………………164

Table 12. Model comparison for predicting N2pc amplitude in mV for sentence Polarity (Affirmative and Negative), Truth value (True and False) and for Adjective position (Shape in nominal position vs colour in nominal position). RD = residual deviance, dAIC = difference between a model's Aikake information criterion and those of the best model, AICw = AIC weight, df = degrees of freedom of the chi-squared statistic, η2 = total eta squared as the ratio between the chi-squared and the residual deviance of the null model……………………………………………………………………………166

Table 13. Model comparison for predicting SPCN amplitude in mV for sentence Polarity (Affirmative and Negative), Truth value (True and False) and for Adjective position (Shape in nominal position vs colour in nominal position). RD = residual deviance, dAIC = difference between a model's Aikake information criterion and those of the best model, AICw = AIC weight, df = degrees of freedom of the chi-squared statistic, η2 = total eta squared as the ratio between the chi-squared and the residual deviance of the null model……………………………………………………………………………………………167

Table 14. Model comparison for predicting RTs for sentence Polarity (Affirmative and Negative), Truth value (True and False) and for Adjective position (Shape in nominal position vs colour in nominal position). RD = residual deviance, dAIC = difference between a model's Aikake information criterion and those of the best model, AICw = AIC weight, df = degrees of freedom of the chi-squared statistic, η2 = total eta squared as the ratio between the chi-squared and the residual deviance of the null model…………………………………………………………………………………………..172

Table 15. Model comparison for predicting N2pc amplitude in mV for sentence Polarity (Affirmative and Negative), Truth value (True and False) and for Adjective position (Shape in nominal position vs colour in nominal position). RD = residual deviance, dAIC = difference between a model's Aikake information criterion and those of the best model, AICw = AIC weight, df = degrees of freedom of the chi-squared statistic, η2 = total eta squared as the ratio between the chi-squared and the residual deviance of the null model…………………………………………………………………………………………..173

Table 16. Model comparison for predicting N2pc amplitude in mV for sentence Polarity (Affirmative and Negative), Truth value (True and False) and for Adjective position (Shape in nominal position vs colour in nominal position). RD = residual deviance, dAIC = difference between a model's Aikake information criterion and those of the best model, AICw = AIC weight, df = degrees of freedom of the chi-squared statistic, η2 = total eta squared as the

ratio between the chi-squared and the residual deviance of the null model………………………………………………………………………………………..………..174

.

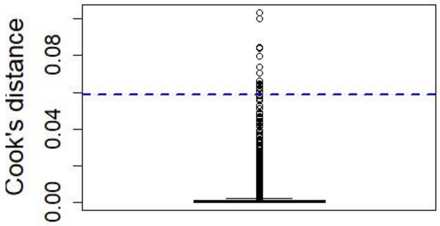
APPENDIX

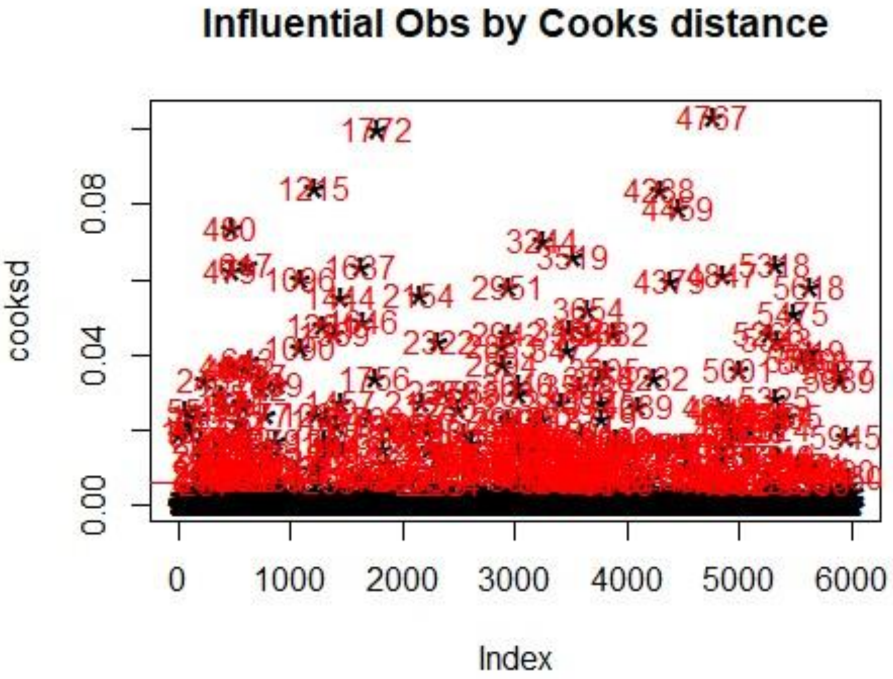
- Outliers were in section 4.1 and 4.2 evaluated by the Influence analysis for Generalized Mixed-Effects models (Nieuwenhuis, Grotenhuis & Pelzer, 2012).
- Audio stimuli of section 4.1 and 4.2
- The ERPs at electrodes PO7/8 contralateral and ipsilateral to the target in all three cue conditions (Polarity, Truth value and Adjective position) of section 4.4
- Random effect plotted fro the section 4.4

Appendix 2 B

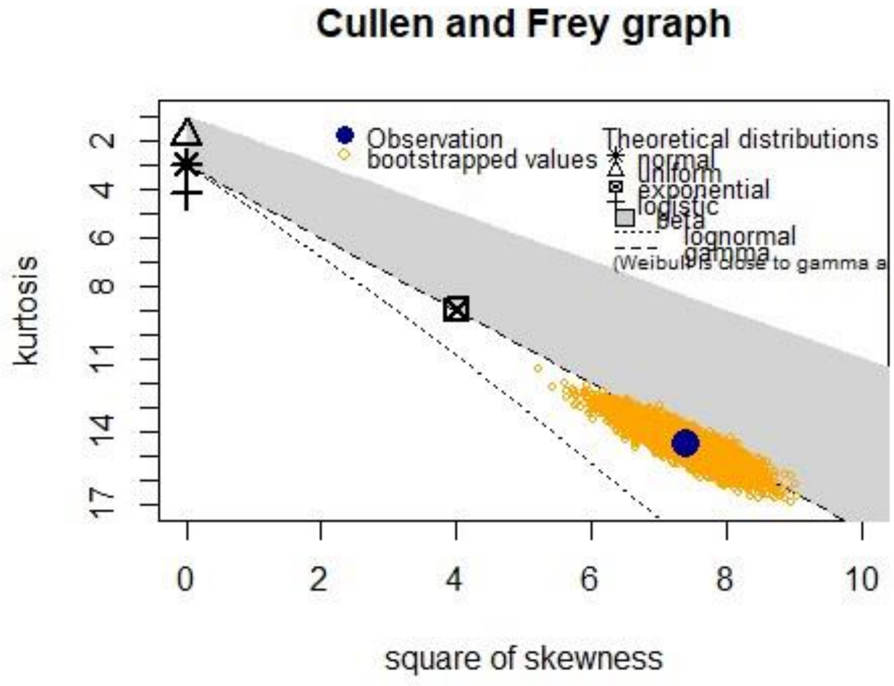
Saccadic latency

Influential analysis of saccadic latency in milliseconds with Cook distance.





Cullen and Frey graph specifying the probability distribution that best fits among predefined family of distributions.`

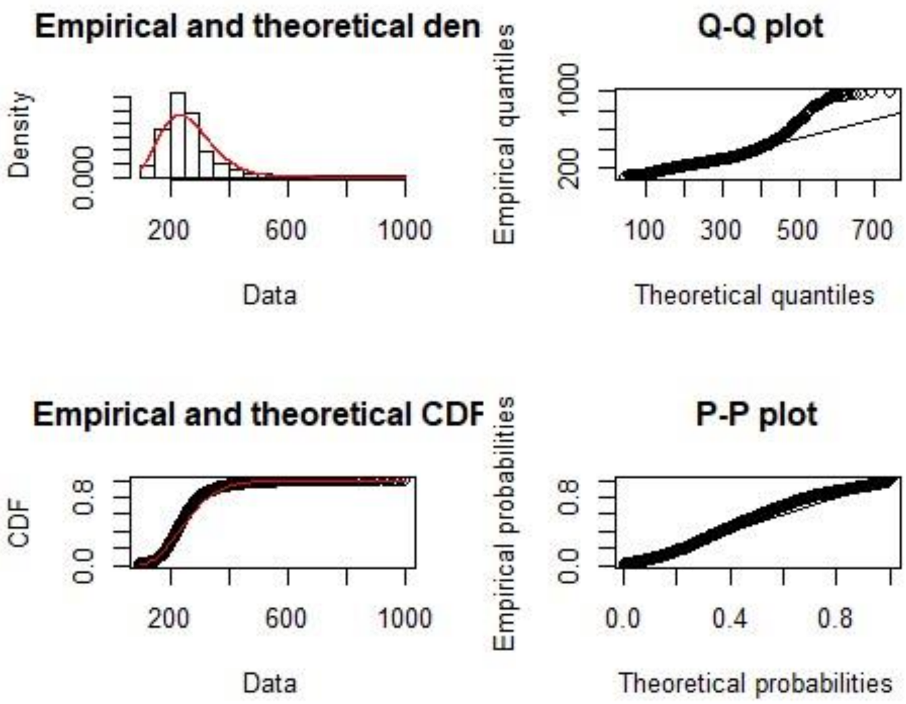


```
## summary statistics ## ------
## min:  100   max:  999
## median:  242
## mean:  263.6615
## estimated sd:  105.0801
## estimated skewness:  2.722167
## estimated kurtosis:  14.52369
```

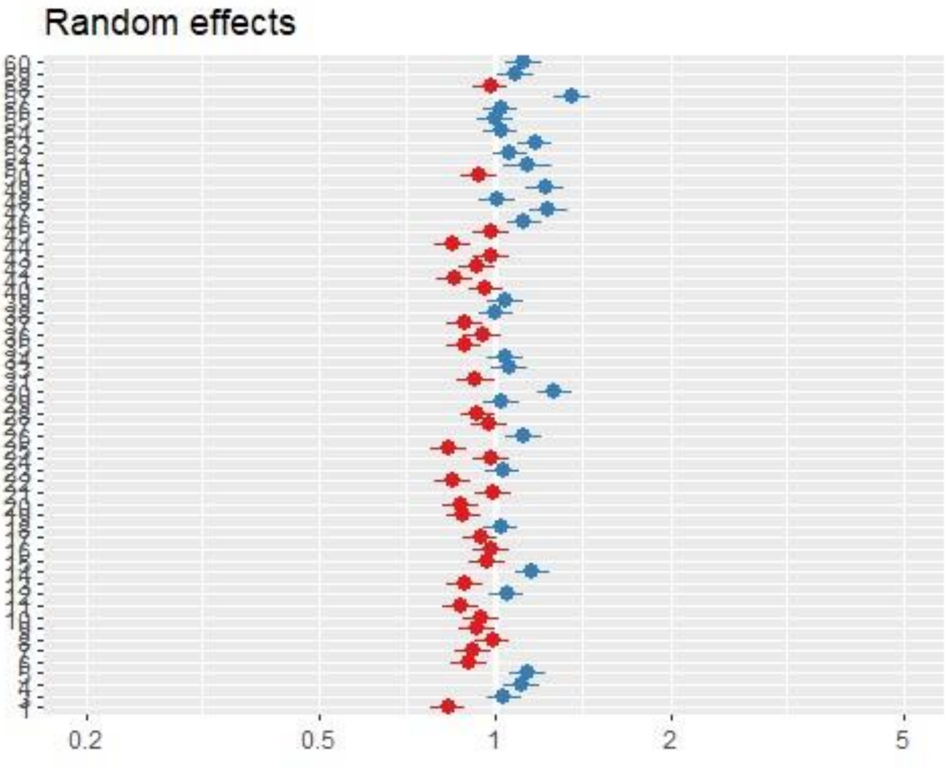
Fit residual with Gamma family with maximum likelihood method (mle)

```
## Fitting of the distribution ' gamma ' by maximum likelihood
## Parameters:
##       estimate   Std. Error
## shape 8.6304867 0.1562400121
## rate  0.0327319 0.0006095497
```

```
## Fitting of the distribution ' gamma ' by maximum likelihood
## Parameters :
##       estimate   Std. Error
## shape 8.6304867 0.1562400121
## rate  0.0327319 0.0006095497
## Loglikelihood:  -33829.71   AIC:  67663.42   BIC:  67676.74
## Correlation matrix:
##          shape      rate
## shape 1.0000000 0.9706045
## rate  0.9706045 1.0000000
```



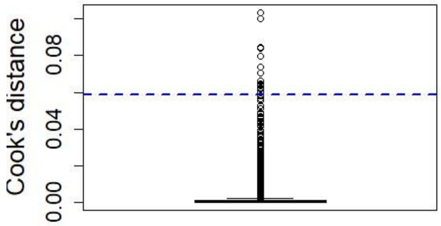
Plot of random intercept of Participants.

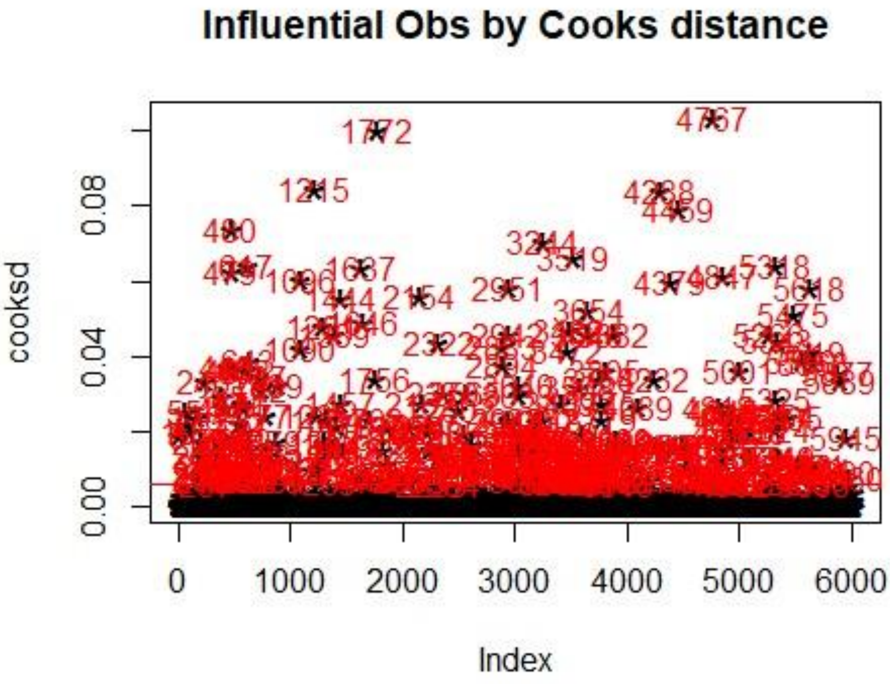


Appendix 2 B

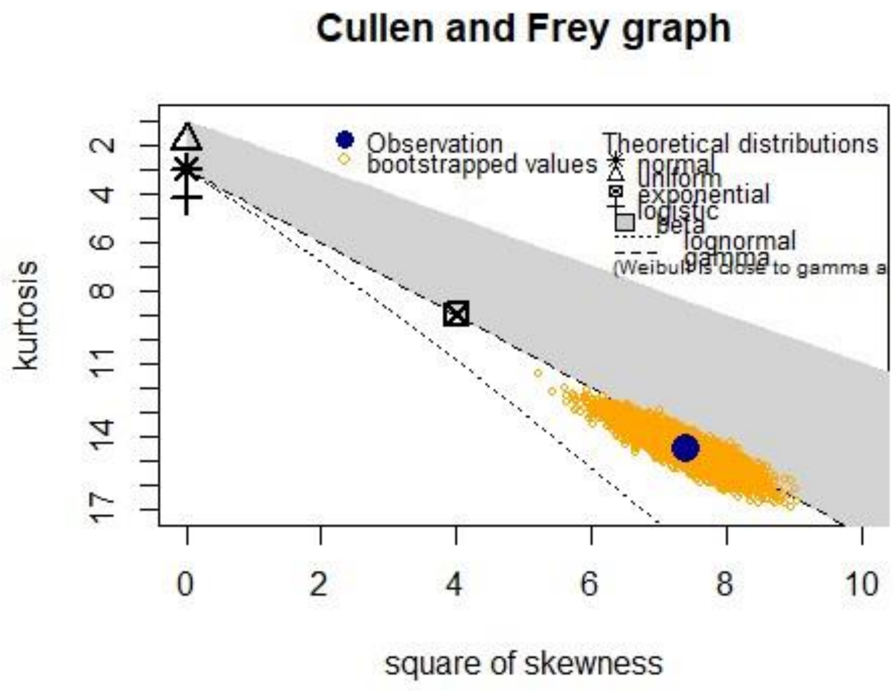
Saccadic latency

Influential analysis of saccadic latency in milliseconds with Cook distance.





Cullen and Frey graph specifying the probability distribution that best fits among predefined family of distributions.`

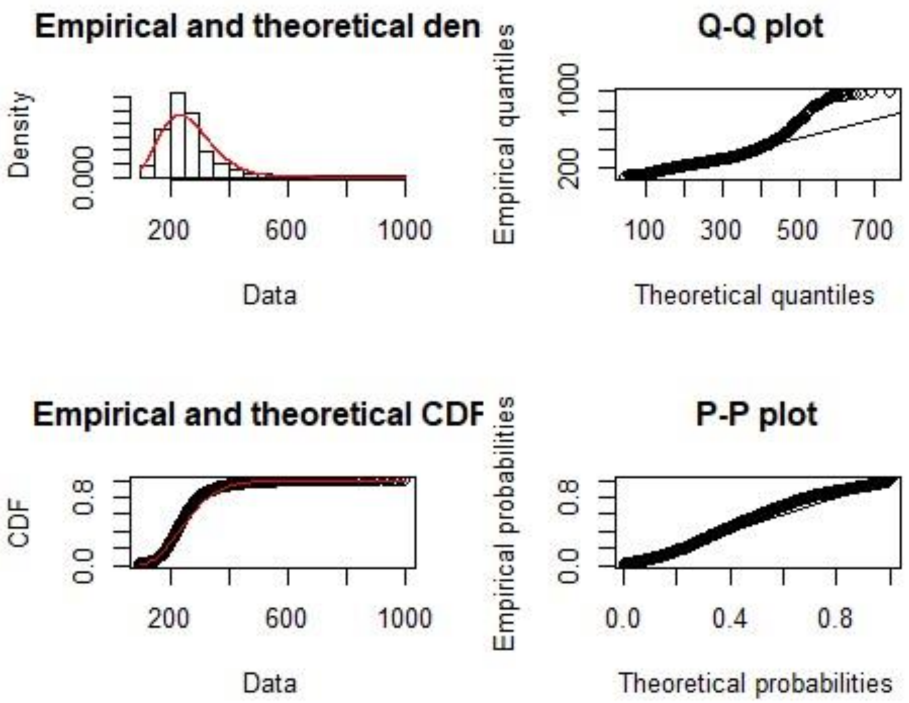


```
## summary statistics ## ------
## min:  100   max:  999
## median:  242
## mean:  263.6615
## estimated sd:  105.0801
## estimated skewness:  2.722167
## estimated kurtosis:  14.52369
```

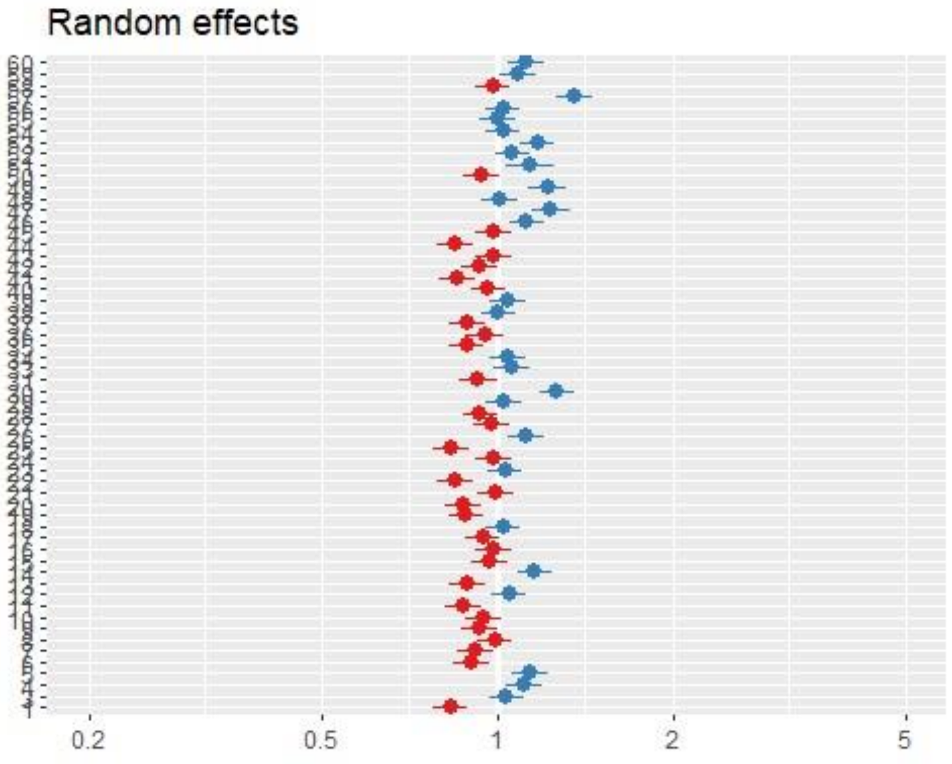
Fit residual with Gamma family with maximum likelihood method (mle)

```
## Fitting of the distribution ' gamma ' by maximum likelihood
## Parameters:
##       estimate   Std. Error
## shape 8.6304867 0.1562400121
## rate  0.0327319 0.0006095497
```

```
## Fitting of the distribution ' gamma ' by maximum likelihood
## Parameters :
##       estimate   Std. Error
## shape 8.6304867 0.1562400121
## rate  0.0327319 0.0006095497
## Loglikelihood:  -33829.71   AIC:  67663.42   BIC:  67676.74
## Correlation matrix:
##          shape      rate
## shape 1.0000000 0.9706045
## rate  0.9706045 1.0000000
```



Plot of random intercept of Participants.



APPENDIX 2

Audio stimuli

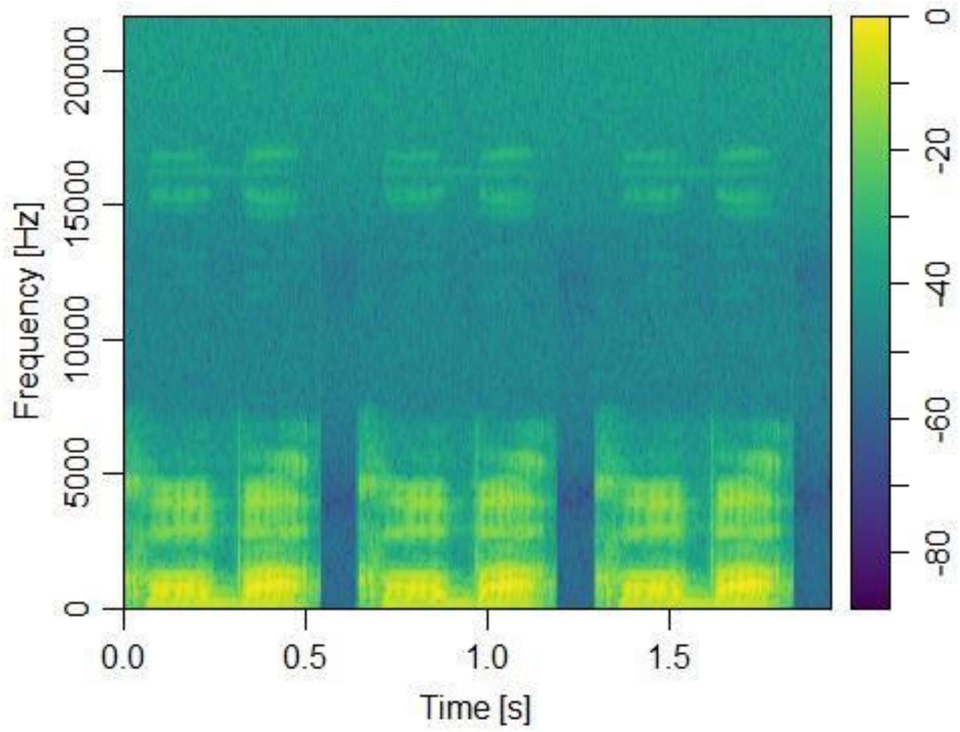
R scprit for audiometric descpritives of pseudo words stimuli from the NOUN database used in Experiment 1 and 2.

```
library(tuneR, warn.conflicts = F, quietly = T)

# read in audio file
data = readWave("coba.wav")

# extract signal
snd = data@left

# determine duration
dur = length(snd)/data@samp.rate
dur # seconds

## [1] 1.951519

# determine sample
rate fs =
data@samp.rate fs #
Hz

## [1] 44100

## [1] 2000

# number of points to use for the fft
nfft=1024

# window size (in points)
window=256

# overlap (in points)
overlap=128

library(signal, warn.conflicts = F, quietly = T) # signal processing functions
library(oce, warn.conflicts = F, quietly = T) # image plotting functions and nice color maps
# create spectrogram
spec = specgram(x =
snd,          n = nfft,
Fs = fs,
         window = window,
overlap = overlap
)

# discard phase information
P = abs(spec$S)
```

```
# normalize
P = P/max(P)

# convert to dB
P = 10*log10(P)

# config time axis t = spec$t


# plot spectrogram imagep(x = t,
      y = spec$f,       z = t(P),
      col = oce.colorsViridis,       ylab = 'Frequency [Hz]',
      xlab = 'Time [s]',       drawPalette = T,       decimate = F
)
```

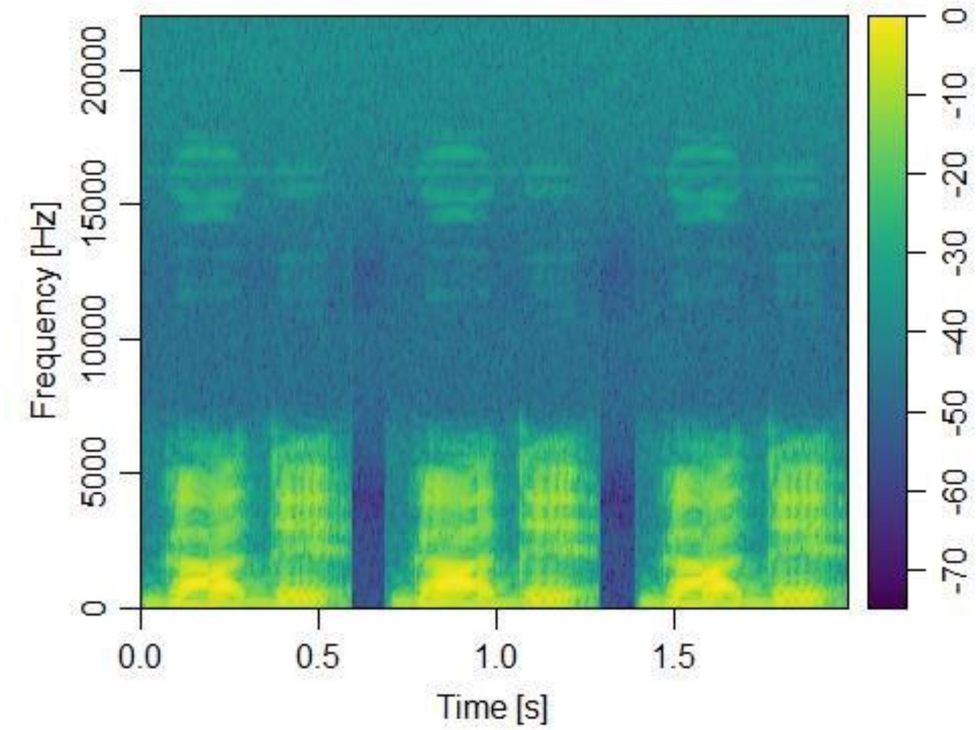
/coba/

```
## [1] 1.951519 ## [1] 44100
```



/gade/

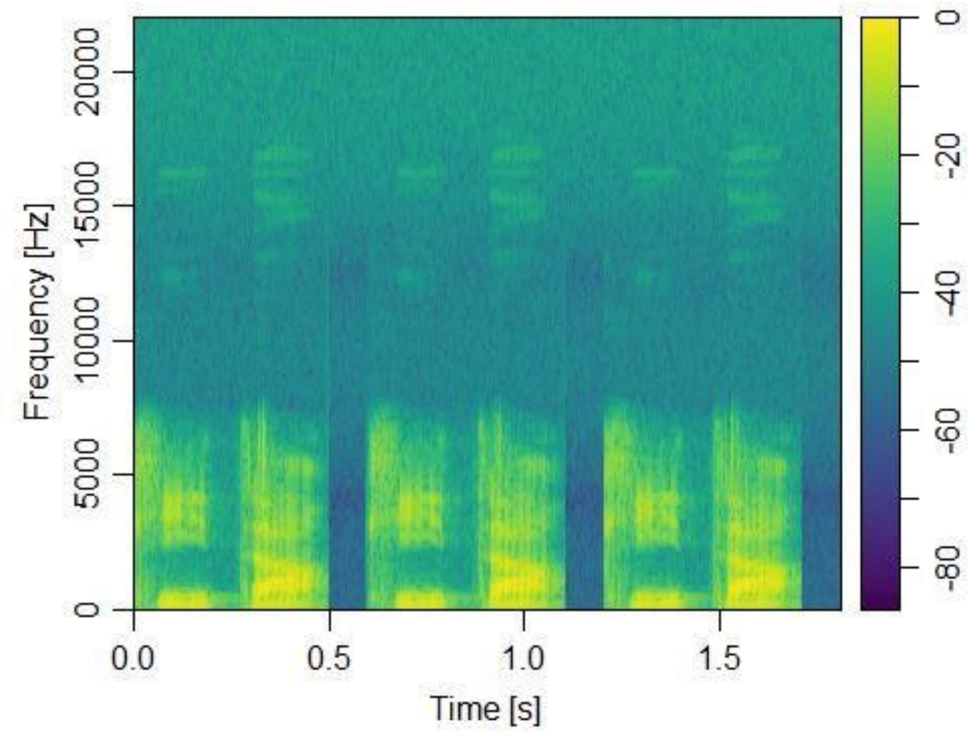
```
## [1] 1.993741  ## [1] 44100
```



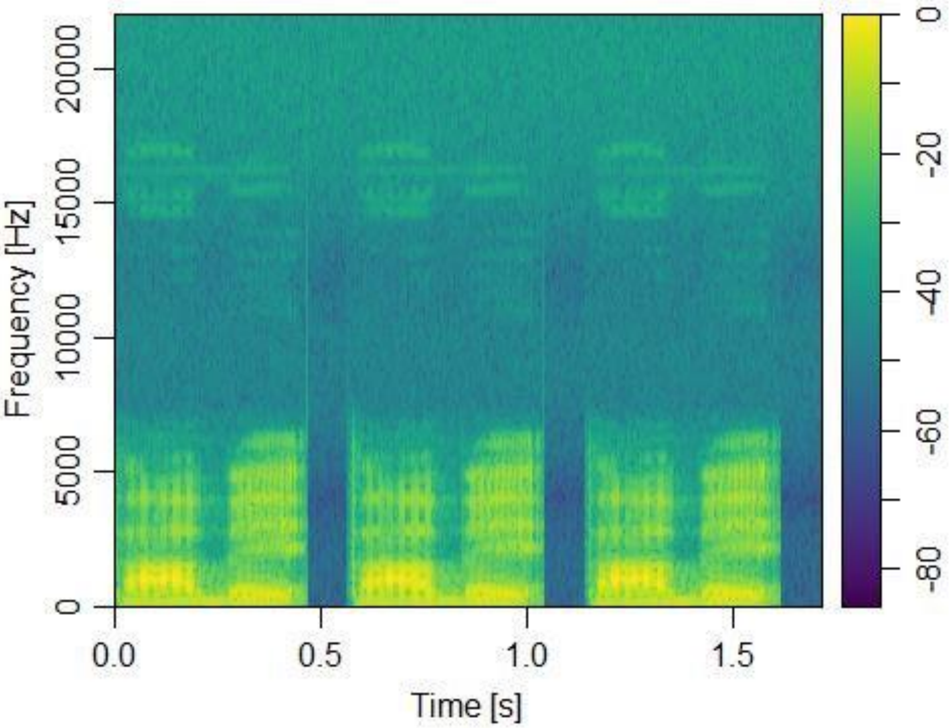
/***kita/***

```
## [1] 1.815102 ## [1] 44100
```

/***pabe/***



```
## [1] 1.71932 ## [1] 44100
```

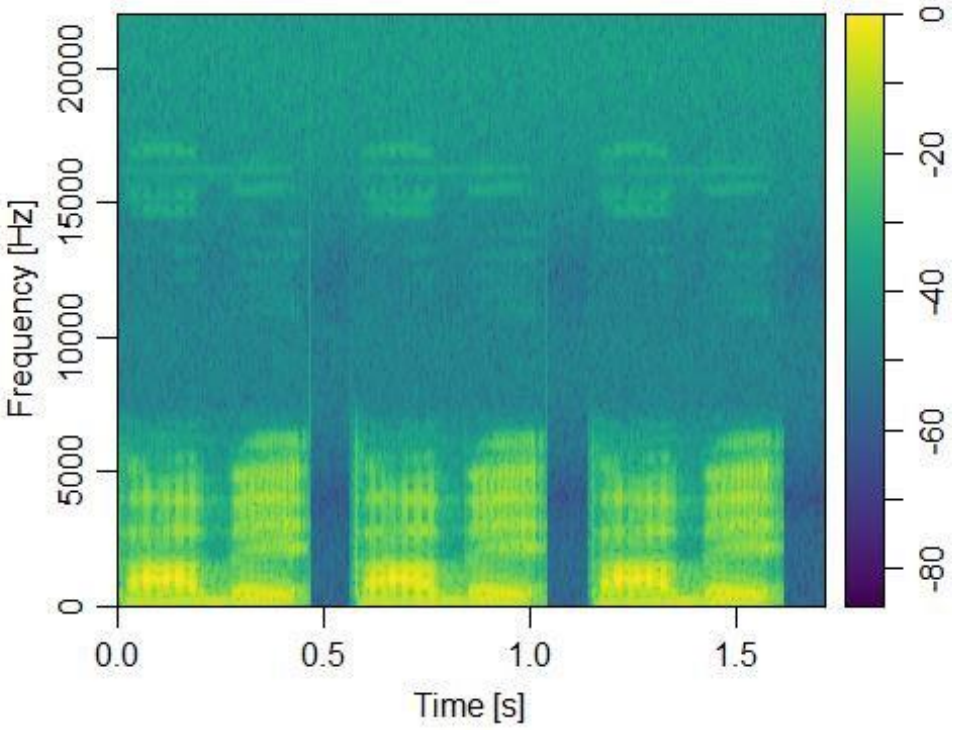


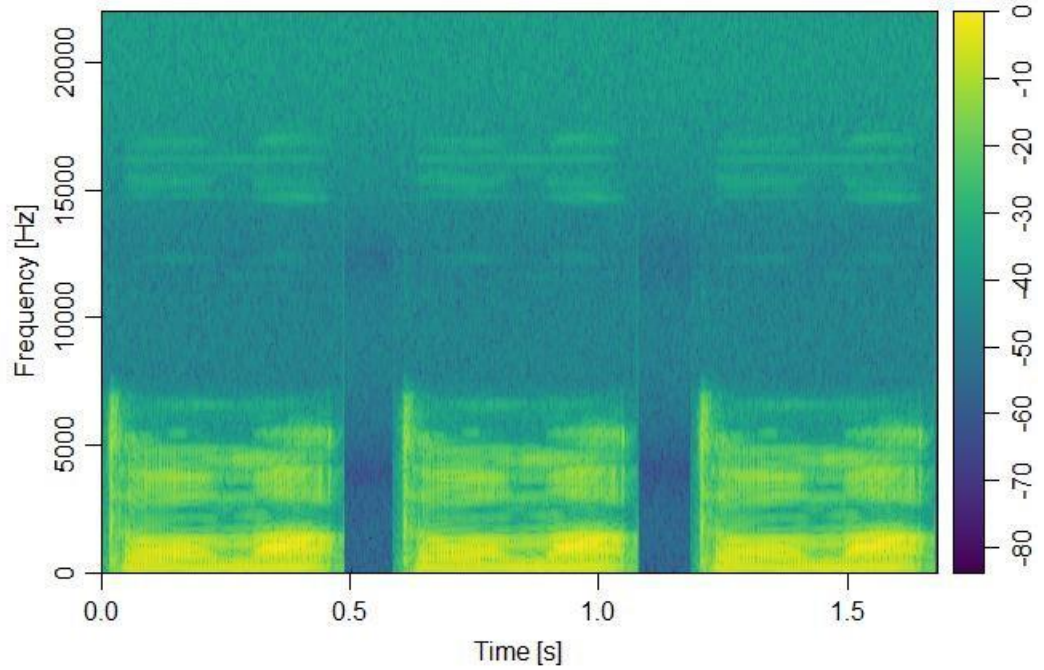
/reda/

```
## [1] 1.815964 ## [1] 44100
```

/toma/

```
## [1] 1. 1.68898  ## [1] 44100
```





APPENDIX 2

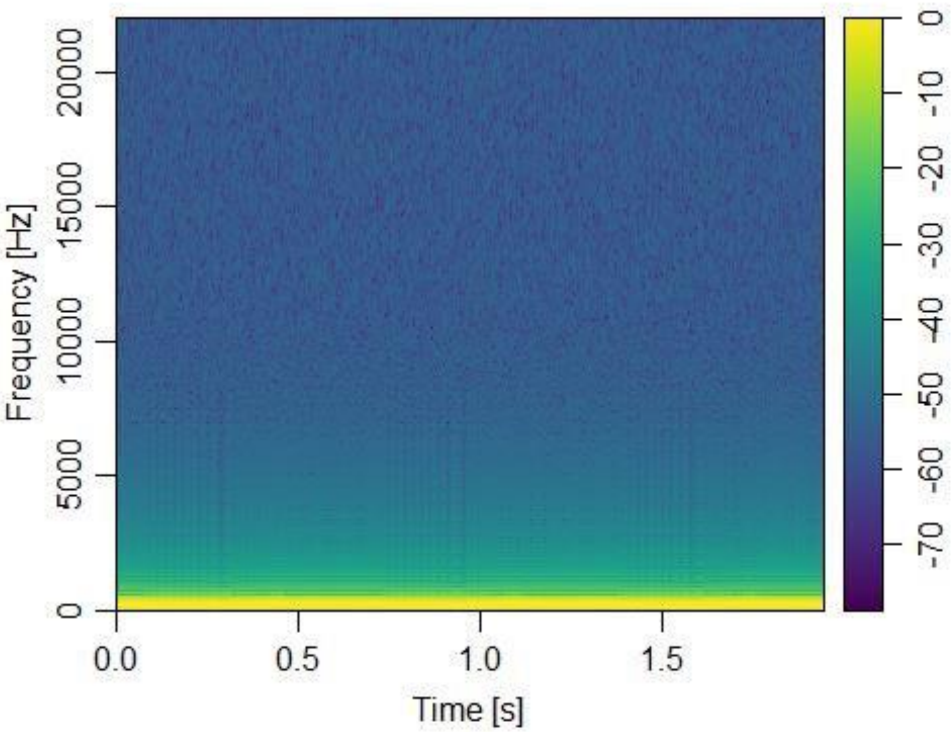
TONES

R scprit and plots for audiometrics descpritives of audio stimuli

/coba/ matched tone

```
# read in audio file data =
readWave("coba_tone.wav")

# extract signal
snd = data@left

# determine duration
dur = length(snd)/data@samp.rate
dur # seconds

## [1] 1.951542

# determine sample
rate fs =
data@samp.rate fs #
Hz

## [1] 44100

## [1] 2000

# number of points to use for the fft
nfft=1024

# window size (in points)
window=256

# overlap (in points)
overlap=128

library(signal, warn.conflicts = F, quietly = T) # signal processing functions
library(oce, warn.conflicts = F, quietly = T) # image plotting functions and nice color maps

## Warning: package 'oce' was built under R version 3.5.3

## Warning: package 'gsw' was built under R version 3.5.3

# create spectrogram
spec = specgram(x =
snd,              n = nfft,
Fs = fs,
              window = window,
overlap = overlap )
```

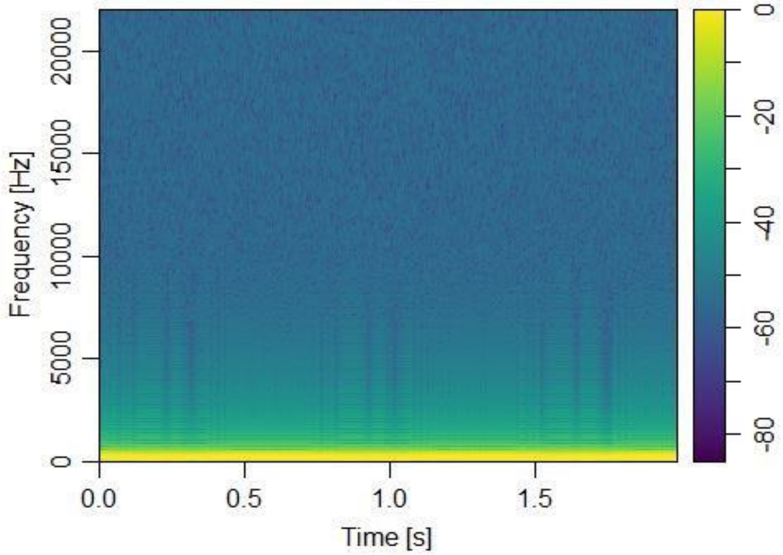
```
# discard phase information
P = abs(spec$S)

# normalize
P = P/max(P)

# convert to dB
P = 10*log10(P)

# config time axis t = spec$t

# plot spectrogram imagep(x = t,
    y = spec$f,     z = t(P),
    col = oce.colorsViridis,     ylab = 'Frequency [Hz]',
    xlab = 'Time [s]',     drawPalette = T,     decimate = F
)
```



/gade/ matched tone

```
## [1] 1.993764
```

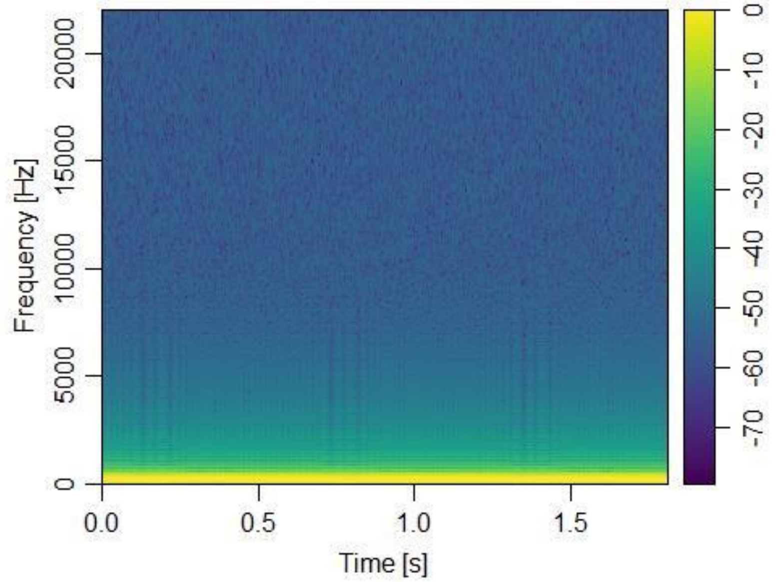
```
## [1] 44100
```



/kita/ matched tone

```
## [1] 1.815125
```

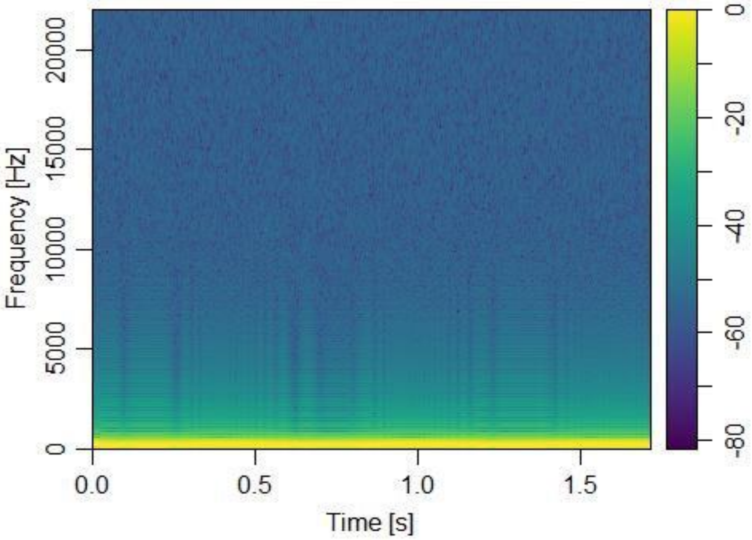
```
## [1] 44100
```



/pabe/ matched tone

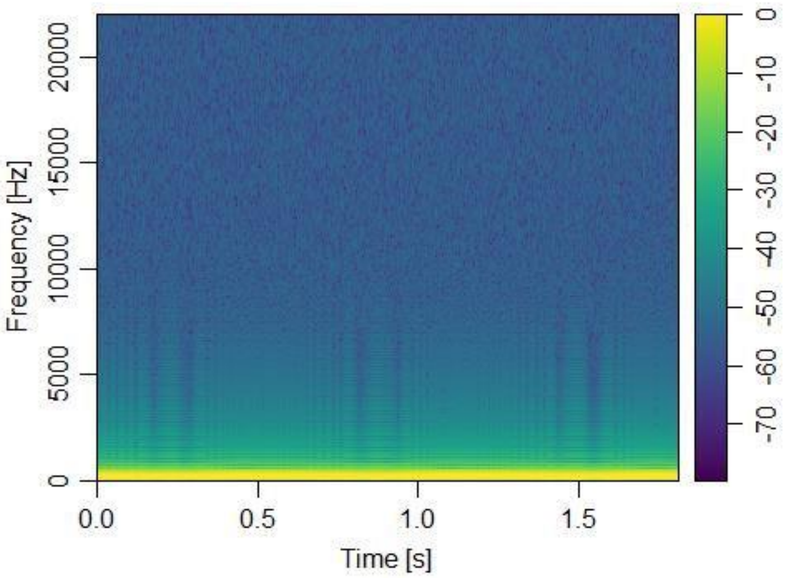
```
## [1] 1.719342
```

```
## [1] 44100
```



/reda/ matched tone

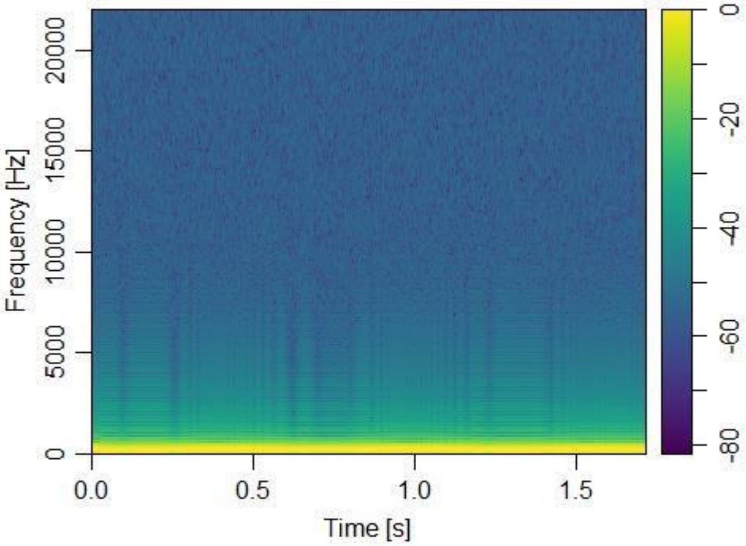
```
## [1] 1.815986
## [1] 44100
```



/toma/matched tone

```
## [1] 1.719342
```

```
## [1] 44100
```



APPENDIX

4.4.

ERPs at electrodes PO7/8 contralateral and ipsilateral to the target in all three cue conditions Truth Value, Polarity and Adjective site, and difference waveforms obtained by subtracting ipsilateral from contralateral ERPs

Plot of Random Effect, i.e. Subjects, and Slopes i.e. Adjective site, Polarity, Laterality and Truth value of n2pc amplitude 180-300ms. For Experiment 1 (up) and 2 (down)